The apoptotic potential of different HIV-1 subtype C Tat mutations in cell culture

Shahieda Isaacs

Thesis presented in partial fulfillment of the requirements for the degree Masters of Sciences in Medical Sciences (Medical Virology) at the Faculty of Medicine and Health Sciences, Stellenbosch University

Supervisor: Prof Susan Engelbrecht

Co-supervisor: Dr Richard Glashoff

Department of Pathology, Division of Medical Virology

March 2013

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Shahieda Isaacs
Name in full

___ 06 __/__ 12 __/__ 2012 _
Date

Abstract

The efficiency in which HIV-1 can infect, spread and evade the attack of therapeutic agents can be attributed to a high mutation rate and frequent recombination events. These factors have collectively contributed to the diversity observed in HIV-1 and resulted in a multitude of subtypes, sub-subtypes, circulating recombinant forms (CRF's) and unique recombinant forms (URF's). The aim of this study was to investigate HIV-1 diversity in Cape Town using a small cohort of treatment naive patients being investigated for HIV Associated Neurocognitive Disorders (HAND). Four different genomic domains: *gag*, *pol*, *accessory* and *gp41* genes were sequenced to subtype the virus. HIV-1 *tat* was further investigated because the dicysteine motif has been reported to play a role in HAND. Viral RNA and proviral DNA was extracted from 64 patients and used for the amplification and sequencing of the genes. Rega and jpHMM online tools were used to identify HIV-1 subtypes and recombinants while Neighbor-joining phylogenetic trees were constructed for phylogenetic analysis. The *pol* gene was further investigated using SCUEAL to detect possible intra-subtype recombination and was also screened for the presence of transmitted drug resistance. In addition *tat* sequence datasets retrieved from the Los Alamos sequence database were investigated and compared with the newly generated sequences for the detection of point mutations and amino acid signature patterns.

Sequencing identified most of the samples as subtype C; however six inter-subtype recombinants (AE, A1G, A1CU and two BC) and 9 intra-subtype C recombinants were identified. In addition 13% of *pol* sequences were identified with resistance mutations. Signature pattern analysis identified a high level of variability in the *tat* sequences: 68% were identified with C30S31; 29% with the C30C31 mutation and a single sequence with a novel mutation C30A31. Functional analysis of these mutations indicated that all mutations investigated were capable of inducing apoptosis in cell culture. The C30C31 mutation generated the highest level of apoptosis, closely followed by the C30A31 mutation. However no statistical significance could be detected between *tat* mutations and the observed levels of apoptosis.

Acknowledgements

My time spent on this project has been riddled with obstacles, but it has without a doubt changed the way I think and approach a problem. I am truly grateful to everyone who has helped me through this project and would therefore like to acknowledge the following individuals and institutions for their contributions and support towards the completion of this thesis:

I am indebted to my supervisor Prof Susan Engelbrecht for providing me with a platform to learn and grow as a young scientist. Her knowledge, guidance and patience have been pivotal in the development of this research project and thesis. I am thankful for have been given the opportunity to work under her supervision.

I am grateful to my co-supervisor Dr Richard Glashoff for his invaluable advice and assistance with regard to the functional analysis of the *tat* mutations. I truly appreciate the time you spent helping me optimize my experiments and for always providing me with a solution when the outcome looked bleak.

I wish to thank Dr Bizhan Romani for his invaluable contribution to this research project. I am grateful for all the advice and time you invested in me and my research. I wish you the best of luck in all your future endeavors.

I wish to acknowledge and thank Mathilda Claassen for all her advice and insight on genetic sequencing and sequence analysis. I am truly grateful for all the help you provided me. To Dalene de Swart, thank you for your constant encouragement, help and all the hilarious stories you've told me over the years which never failed to brighten up my day. To Randall Fisher, thank you for all the advice and assistance you provided me whenever problems arose, your support was greatly appreciated.

To Dr Graeme Jacobs, thank you for your assistance in the laboratory as well as taking the time to review and critically assess my thesis. Your continued support, criticisms and thoughts have definitely helped to shape my thesis into something I can be proud of. Thank you to Nafiisah Chotun, my writing buddy in GERGA, your words of encouragement kept me motivated while writing this thesis.

I would like to acknowledge the support of Prof John Joska (UCT) and Prof Soraya Seedat for providing me with the samples for this study. Without which this research could not have been possible.

I am thankful to the National Health Laboratory Services Research Trust (NHLS RT) and the Poliomyelitis Research Foundation (PRF) for their financial assistance as well as the NIH AIDS Research and Reference Reagent Program for providing me with the necessary reagents to continue functional analysis of HIV-1 Tat.

To all the staff in the Division of Medical Virology, thank you for your moral support throughout the duration of my master's degree. To My fellow students, especially, Germaine, Heleen, Karmistha, Marilize, Nafiisah, Ndapewa and Rozanne, I am grateful that I was blessed to have met and befriend you. My years spent in this department have been good but it was even more enjoyable and memorable because of you. I will forever cherish your friendship and support.

Finally I wish to thank my family for their constant support and encouragement throughout my academic career.

Presentations at Meetings

Ms S Isaacs, Dr R Glashoff, Prof S Seedat, Dr J Joska, Dr Robert Paul and Prof S Engelbrecht. Molecular Characterization of Tat: A Possible mediator of HIV-1 Associated Dementia (HAD). Virology Africa. Cape Town, South Africa, 29 November – 2 December 2011. (Poster).

Ms S Isaacs, Dr R Glashoff, Prof S Seedat, Prof J Joska, Dr Robert Paul and Prof S Engelbrecht. Molecular Characterization of HIV-1 Tat. 56th Annual Academic Year Day, Tygerberg, South Africa, 17 -18 August 2011. (Presentation).

Biology is the science. Evolution is the concept that makes biology unique.

Jared Mason Diamond

American ***Scientist***

List of Abbreviations

A	Absorbance
AIDS	Acquired Immunodeficiency Syndrome
AA	Amino acid
ANOVA	Analysis of variance
ART	Anti-retroviral therapies
ANI	Asymptomatic neurocognitive impairment
bFGF	Basic fibroblast growth factor
BBB	Blood Brain Barrier
BSA	Bovine serum albumin
CA	Capsid protein
CNS	Central nervous system
CTD	Carboxyl-terminal domain
CSF	Cerebrospinal fluid
cDNA	Complementary deoxyribonucleic acid
CPX	Complex
CRFs	Circulating recombinant forms
CDK7	Cyclin-dependant kinase 7
DTT	Dithiothreitol
ddNTP's	Dideoxyribonucleoside triphosphates
dNTP's	Deoxyribonucleoside triphosphates
DMEM	Dulbecco's modified Eagle's medium
EDTA	Ethylenediamine-tetraacetic acid
Exo	Exonuclease 1
ExoSapIT	Exonuclease Shrimp Alkaline Phosphate
Env	Envelope
FBS	Fetal bovine serum
Gag	group-specific antigen
HAART	Highly Active Anti-retroviral Therapy
HIV	Human immunodeficiency virus
HAND	HIV Associated Neurocognitive Disorders
IN	Integrase
IDT	Integrated DNA Technologies
IFN	Interferon gamma

IL-1β	Interleukin-1 beta
Kb	Kilobase
miRNA	micro RNA
MNGC	Multinucleated Giant Cells
NF-KB	Nuclear factor-kappa Beta
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside transcriptase inhibitor
PBS	Phosphate buffered saline
PCP	*Pneumocystic carinii* pneumonia
PCR	Polymerase chain reaction
Pol	Polymerase gene
PIC	Pre-integration complex
PI	Protease inhibitor
PR	Protease
ROS	Reactive oxygen species
RT	Reverse transcriptase
RRE	Rev responsive element
RTC	Reverse transcription complex
RNA	Ribonucleic acid
SQV	Saquinovir
SAP	Shrimp Alkaline Phosphate
TAK	Tat-associated kinase
Taq	*Thermus aquaticus*
Tfl	*Thermus flavus*
TNF-α	Tumor necrosis factor alpha
TAR	Trans-activation responsive element
tat	Transcriptional transactivator gene
Tat	Transcriptional transactivator protein
TRAIL	Tumor necrosis factor-related apoptosis-inducing ligand
WT	Wild-type

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

In 2011, it was estimated that 23.5 million people were living with HIV-1 in sub-Saharan Africa with an adult prevalence of between 15-49% (UNAIDS, 2012). Although the HIV epidemic has reached a plateau in South Africa, the country still has the highest prevalence of HIV infected individuals. However, an increase in prevalence has been reported in individuals aged 25 years and above as a result of the availability of ARV treatment and elevated levels of promiscuity. The prevalence of HIV in the different provinces of South Africa is variable, with the highest prevalence of 39.5% (2009-2010), among 15-49 year olds, recorded in KwaZulu-Natal. Other provinces with high prevalence levels include Mpumalanga (35.1%), Free State (30.6%) and Gauteng (30.4%). A significant increase in prevalence of 16.9% to 18.5% was also observed in the Western Cape in 2010. A recent antenatal survey also indicated that most of these HIV-1 infections occurred in young women aged between 30-34 years (Global AIDS Progress Report, 2012). Excessive genetic variability is a main feature of HIV-1 epidemiology and has resulted in 9 subtypes, 2 sub-subtypes, at least 55 circulating recombinant forms (CRFs) and numerous unique recombinant forms (URFs) being identified (http://www.hiv.lanl.gov).

Epidemiological studies have identified subtype C as the most prevalent form of HIV-1 in the world (Ariën et al., 2007). Subtype C dominates in regions of sub-Saharan Africa, India, Brazil and China and is responsible for almost 50% of all global HIV-1 infections. Subtype B circulates in the Americas, Western Europe and Australia while subtype A is the main HIV-1 circulating form in Eastern Africa (Kenya, Uganda, and Tanzania) and Eastern Europe. Due to recombination and high mutation rates in the life cycle of HIV-1, viral subtypes can differ by as much as 20-25% and therefore have varying degrees of pathogenicity, disease progression, co-receptor usage and transmissibility (Ataher et al., 2012).

The aim of this study was to subtype and characterize HIV-1 in a cohort of patients being investigated for HAND. In the literature review, I will briefly discuss the characteristics of HIV-1, the diversity of the virus and focus on the structure and function of the HIV-1 *tat* gene and Tat protein. HIV-1 and neurocognitive disorders will also be briefly reviewed. Finally, the aim and objectives of the study will be stated.

LITERATURE REVIEW

1.1. Human Immunodeficiency Virus type-1

1.1.1 The history and epidemiology of HIV-1

AIDS was first discovered in the 1980s when five men presented with *Pneumocystis carinii* pneumonia (PCP), a rare opportunistic infection that affected individuals with compromised immune systems (Follansbee et al., 1982). These men died within months and soon more men presented with a rare skin cancer called Kaposi sarcoma (Gottlieb et al., 1981). In both instances, the underlying cause for the infection was unknown. However these infections were not isolated and more cases arose which prompted the monitoring of the outbreak. Similar cases were also reported in Africa, but manifested in both homosexual and heterosexual populations (Piot et al., 1984). A disease pattern was soon identified and was characterized by a low $CD4^+$ cell count, lymphadenopathy and infection by opportunistic infections (Basavapathruni & Anderson, 2007).

In 1983 Montagniers' group isolated a virus named "lymphadenopathy-associated virus," or LAV from a patient with persistent lymphadenopathy (Barré-Sinoussi et al., 1983). The virus appeared to be related to the *Lentivirus* genus of the *Retroviridae* family and was named Human Immunodeficiency Virus (HIV) (Levy et al., 1984; Basavapathruni & Anderson, 2007). HIV-1 has since spread across the world, but in an irregular distribution of subtypes resulting in HIV-1 hotspots like sub-Saharan Africa. The global spread of HIV-1 and the areas dominated by specific subtypes are indicated in **Figure 1.1.**

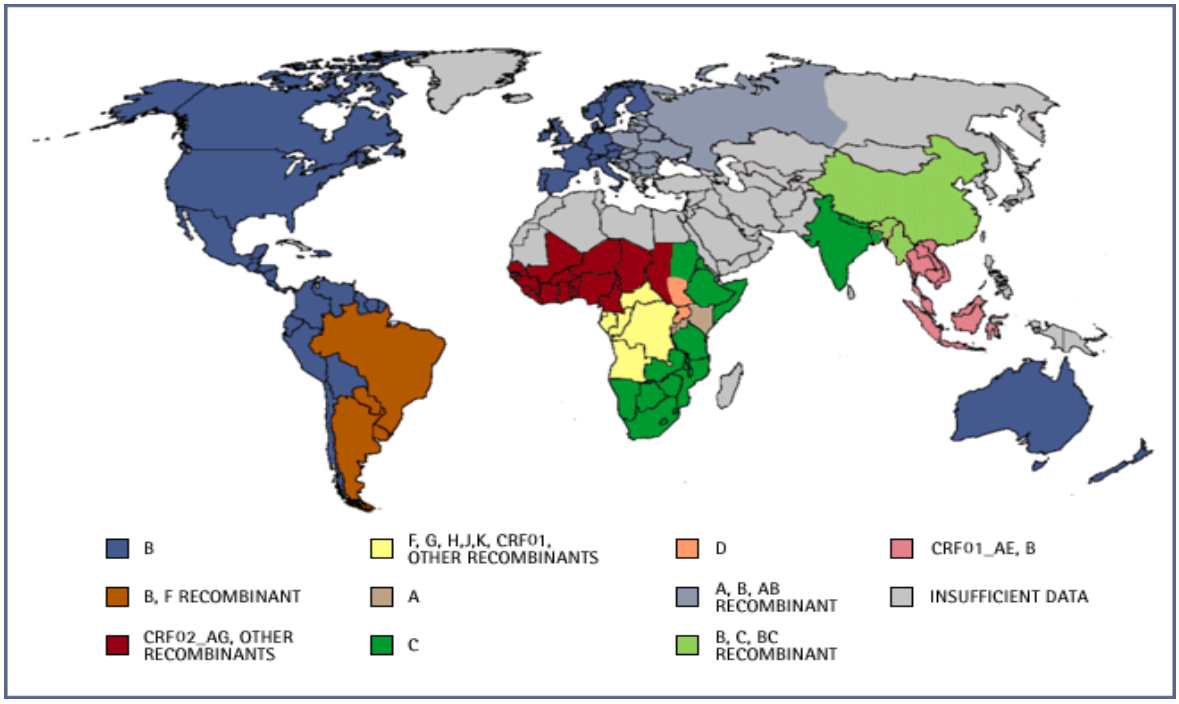

Figure 1.1. Circulation of HIV-1 subtypes and recombinant forms in the world. This geographical distribution highlights the prevalence of subtype C in southern Africa and Subtype B in economically developed regions such as North America, Australia and Europe. (Source: http://www.pbs.org/wgbh/pages/frontline/aids/art/clademap.gif

1.1.2 HIV-1 viral structure and gene organization

HIV-1 belongs to the Retroviridae family and *Lentivirus* genus and is approximately 100nm in diameter. Its genome consists of two copies of positive stranded viral RNA encased in a capsid which is surrounded by a lipid bilayer envelope (*env*) with surface glycoprotein spikes (**Figure 1.2)**. The glycoproteins spikes are derived from the *env* gp160 precursor and composed of the outer surface protein (gp120) and the transmembrane protein (gp41), which anchors the spike to the lipid bilayer (Briggs et al., 2003; Levy, 2007).

The matrix protein (p17) lines the inner surface of the HIV-1 virion and encases the icosahedral shell of the p24 capsid protein (CA). and the viral enzymes *protease* (PR), *integrase* (IN), and *reverse transcriptase* (RT) (Nisole & Saïb, 2004). *Pol* is responsible for the encoding of these enzymes *protease*, *reverse transcriptase, RNase H* and *integrase.* Reverse transcriptase is necessary for the transcription of

viral RNA into cDNA while integrase is responsible for the integration of the viral DNA into the host genome (Levy, 2007; Dicker et al., 2007). The remaining regulatory and accessory genes control viral replication and HIV's ability to infect cells (Chan et al., 1997).

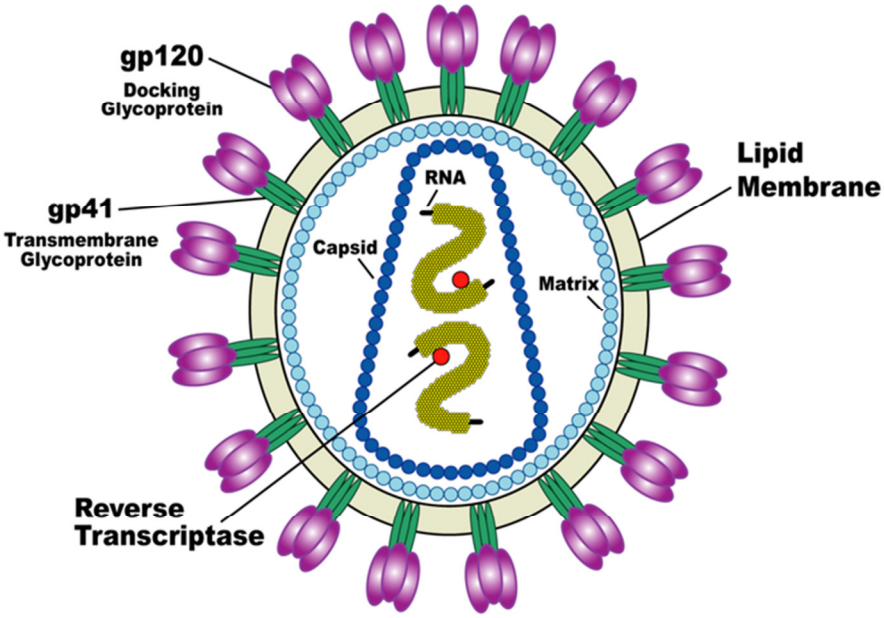

Figure 1.2 Graphical presentation of the HIV-1 virion structure. The diagram shows the lipid bilayer containing the surface glycoprotein spikes (gp120 and gp41), the matrix protein (p17) which lines the inner surface of the HIV-1 virion and the two copies of RNA. Reverse transcriptase integrase and protease enzymes can be observed within the p24 capsid, and are necessary for the reverse transcription of viral RNA, integration of proviral DNA into the host genome and the cleaving of mature polyproteins (Source: http://en.wikipedia.org/wiki/File:HIV_Virion-en.png and http://web.archive.org/web/20041119131214/http://www.niaid.nih.gov/factsheets/graphics/howhiv.jpg)

The HIV-1 genome is approximately 9.2kb long and composed of structural landmarks (LTR, tar, rre, pe, slip, crs and INS), structural genes (*gag*, *pol* and *env*), regulatory genes (*tat* and *rev*) and accessory genes (*nef*, *vif*, *vpr* and *vpu*) as indicated in **Figure 1.3** (HIV sequence compendium, 2011). The long terminal repeats (LTRs) flank the genome of the integrated provirus and regulate the production of new virus particles with the aid of either HIV-1 or the host cell proteins. The U3 region of the LTR contains elements which recruit transcription factors and

initiates transcription of viral RNA. U3 transcription activity is low but can be enhanced through regulatory proteins such as Tat (Suzuki & Suzuki 2005). The structural genes *gag*, *pol* and *env* codes for proteins essential for viral structure. The Rev responsive element (RRE) encoded within the *env* gene facilitates protein binding. The SLIP (TTTTTT) element is necessary to generate functional *pol*. The *gag* gene is a necessary structural component of the virus and generates the viral capsid (p24), the nucleocapsid (p6 and p7) and the matrix (p17) (HIV Sequence Compendium 2011).

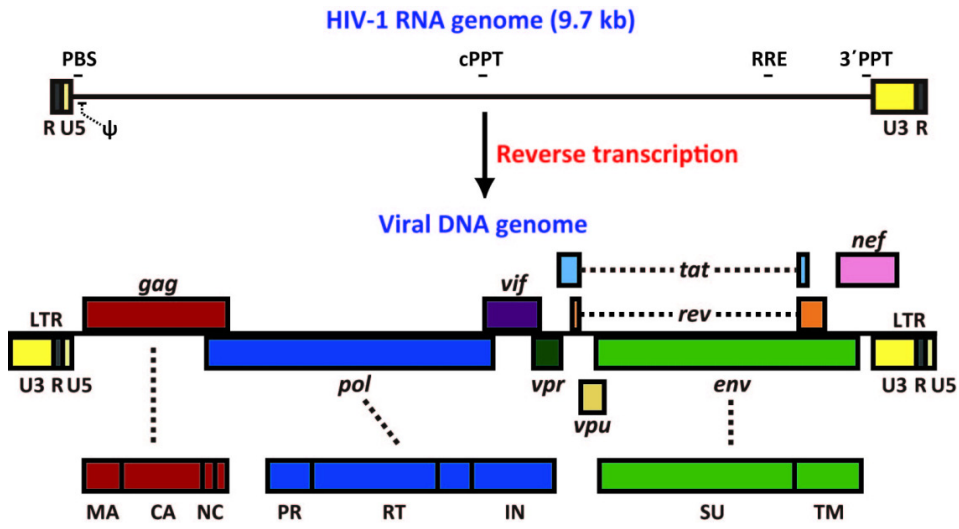

Figure 1.3: Genome organization of HIV-1. Viral RNA is transcribed into DNA comprised of LTRs flanked on either ends of the genome. Structural genes (*gag*, *pol* and *env*) regulatory genes (*tat* and *rev*) and accessory genes (*nef*, *vif*, *vpr* and *vpu*) are indicated as well as the proteins encoded by each genomic region (Source: Suzuki & Suzuki, 2005), Reprinted with permission from Prof Yuichi Suzuki.

The RNA response element TAR is a 59 base stem loop structure localized within the first 80 nucleotides of the LTR. Deletion studies have shown that nucleotides 19 to 42 are sufficient for Tat binding. The TAR RNA contains a 6- nucleotide loop and a three-nucleotide pyrimidine bulge which separates helical stems and plays a critical role in HIV replication by providing a binding site for the recruitment of the viral

transactivator protein Tat (Lalonde et al., 2011). In addition, it has been proposed that TAR plays other roles in replication such as: dimerization, strand transfer during reverse transcription, a possible HIV-1-derived microRNA (miRNA) during latency, and viral packaging (Lu et al., 2011). Mutational studies from Das et al., (1998) concluded that the disruption of TAR resulted in reduced levels of genome packaging and this was supported by similar results (Clever & Parslow, 1997). A summary of HIV-1 protein functions is listed in **Table 1.1**.

Table 1.1 HIV-1 proteins and their function (Source: Frankel & Young, 1998).

Protein	**Function**	**Abbreviation**
Gag	Capsid structural protein	CA
	Matrix protein, myristoylated	MA
	Nucleocapsid protein, helps in reverse transcription	NC
	Role in viral budding (L domain)	-
Polymerase	Reverse Transcription (RT): RNase H-inside core	RT
Protease	Gag/Pol cleavage and maturation	PR
Integrase	Viral cDNA integration	IN
Env	Envelope surface protein	SU
	Envelope transmembrane protein	TM
Tat	Transactivation	Tat
Rev	Regulation of viral mRNA expression	Rev
Nef	Pleiotropic, can increase or decrease virus replication	Nef
Vif	Promotes virion maturation and infectivity	Vif
Vpr	Helps in virus replication, transactivation	Vpr
Vpu	Helps in virus release, disrupts gp160:CD4 complex	Vpu
Vpx	Helps in entry and infectivity (HIV-2 and SIV)	Vpx

1.1.3 The life cycle of HIV-1

The life cycle of HIV-1 has two distinct stages, the early phase and the late phase. The early phase entails cell binding and integration of the transcribed viral DNA into the cell genome, while the late phase begins with the expression of viral genes and ends with the release of mature progeny virions (D'Souza & Summers, 2005). The initial step of the early phase is the binding and adsorption of viral particles to the cell surface proteins of the viral lipid bilayer. Interactions between viral envelope glycoproteins and $CD4^+$ receptors on host T-cells facilitate viral entry into the host cell. This interaction leads to conformational changes in the $CD4^+$ T cell receptor and gp120 glycoproteins. In addition to this, the recruitment of the chemokine co-receptors CXCR4 and CCR5 is initiated. A second interaction between CXCR4 or CCR5 and gp120 results in a second conformational change. These conformational changes culminate in the dissociation of gp120 from gp41, and the alteration of gp41 to its fusogenic conformation. Entry of virions into host cells is finally achieved by the insertion of gp41 fusion peptide into the target membrane as illustrated by **Figure 1.4.**

Fusion of viral and cellular membranes releases the viral genome into the cytoplasm. The viral core is uncoated and forms a pre-integration complex (PIC). The PIC is translocated across the nuclear pore during cell division and into the nucleus. Early phase infection is completed once the viral genome has been integrated into the host genome using the viral protein integrase (Nisole & Saïb, 2004). Infected resting memory $CD4^+$T cells serves as a source of latent infection and supports continued HIV-1 vireamia (Kim & Perelson, 2006). The late phase is initiated by the expression and release of viral proteins from the nucleus and ends with the formation of new progeny virus. Once integrated, the body recognizes the foreign DNA as its own and begins the process of transcribing viral RNA. Spliced and unspliced mRNA transcripts are translocated across the nuclear pore and translated within the cytoplasm. Translated gag polyproteins are recognized by other viral and cellular components and collectively transported to the cell membrane where it is packaged into new virus progeny and released from the host cell by budding as immature viral progeny (Briggs et al., 2003; Freed, 2001). During or after the budding process, protease will induce structural changes and viral maturation via the proteolytic

processing of gag p55. Continuous immune activation results in host cell death via apoptosis, causing a severe loss of $CD4^+$ T-cells (Roshal et al., 2001; (Badley et al., 2003; Badley, 2005).

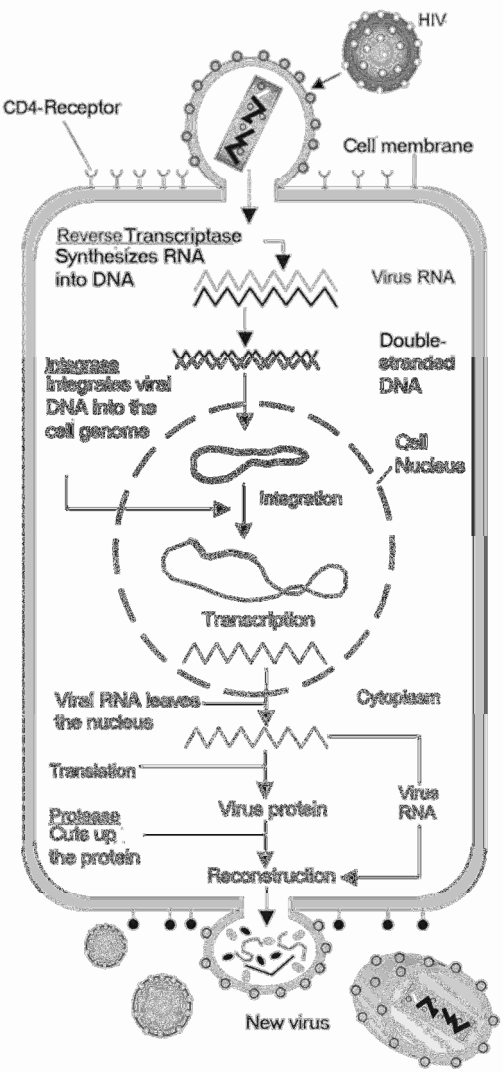

Figure 1.4 HIV-1 viral life cycle. HIV-1 virus binds and fuses to host cell, fusion facilitates release of viral RNA and reverse transcription is initiated. Viral DNA translocates into nucleus and is integrated into host genome. The host cell transcribes viral DNA and immature viral proteins are assembled and released from host cell, where the virion develops into mature virus. (Source: http://upload.wikimedia.org/wikipedia/commons/3/35/HIV_gross_cycle_only.png).

1.1.4 Classification of HIV-1

HIV can be divided into two types, HIV-1 and HIV-2. Phylogenetic analysis has been used to classify the virus and also identify the distribution pattern of these strains

across the world. HIV-2 originated from SIV/SMM infected sooty mangabeys (*Cercocebus atys*). It is exclusive to West and Central Africa. HIV-1 on the other hand can be found in various regions of the world. HIV-1 is further divided into four groups, M (major), O (outliers) and N (Non-M, non-O) which originated from SIV infected chimpanzees (*Pan troglodytes troglodytes*) (Gao et al., 1999). In addition a newly discovered group P that originated from gorillas has been described (Plantier et al., 2009). The M group is responsible for most of the world's HIV infection and can be divided into nine different subtypes (A-D, F-H, J and K) according to *env* gene diversity (**Figure 1.5**). Epidemiological studies have suggested that these subtypes each evolved from numerous founder viruses during epidemics in Africa and that this continued evolution eventually produced sub-subtypes and at least 55 Circulating Recombinant Forms (CRF's) (http://www.hiv.lanl.gov). CRF's are identified according to a numerical system and letters which present their parental subtype (Ball et al., 2003; Tebit & Arts, 2011). In areas of subtype co-circulation, up to 10% of HIV-1 infections are a consequence of unique recombinant forms (URFs) (Pond and Smith 2009). A number of factors, such as dual HIV-1 infection, promote the emergence of URFs and unlike the slow accumulation of mutations through replication errors, this form of recombination causes rapid sequence diversity (Ragupathy et al., 20011).

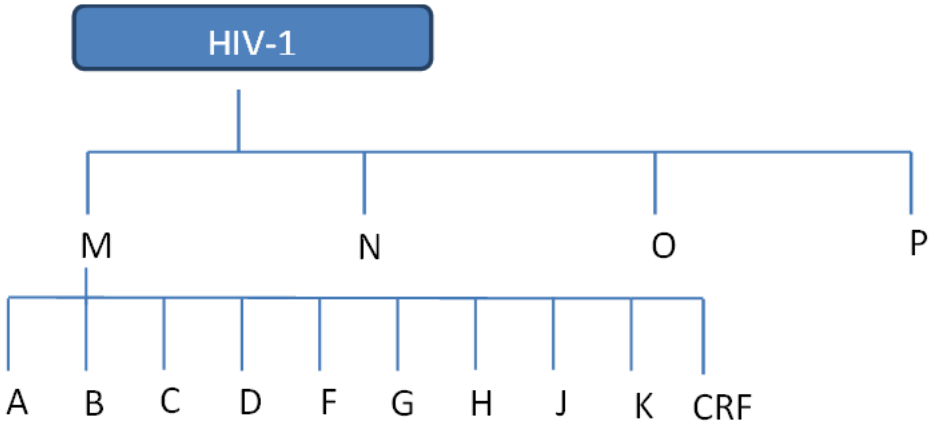

Figure 1.5 Classification of HIV-1. The classification of HIV and the various groups, subtypes and CRF's of HIV-1 are indicated in the diagram. HIV-1 is shown to have 4 groups: M,N, O and P with the group M sub-dived further into subtypes A, B, C, D, F, G, H, J, K and CRF's.

1.2. HIV-1 diversity in South Africa

In 1982 the first case of HIV-1 was documented in South Africa (Ras et al., 1983; Becker et al., 1985). The reported cases initially seemed to be localized amongst homosexual white men and were dominated by subtype B and subtype D. However by the beginning of the 1990's heterosexual members of the indigenous African population were also infected by HIV-1 (Sher, 1989; Engelbrecht et al., 1995; Williamson et al., 1995). Heterosexual infections has since become the prevalent form of HIV-1 transmission, with subtype C as the main subtype circulating in infected South Africans. However, a number of non-subtype C viruses and recombinants have also been reported (Engelbrecht et al., 1995; Van Harmelen et al., 1997; Bredell et al., 2002; Papathanasopoulos et al., 2002; Loxton et al., 2005; Rousseau et al., 2006). A study by Jacobs et al., (2009) identified their cohort sequences as predominantly subtype C (89%) followed by subtype B (6.8%), subtype A (3.1%) non-subtype C's (F1, G, U) and a CH recombinant. Another study which investigated the HIV-1 diversity of Cape Town between1984 and 2010 showed that earlier HIV-1 sequences consisted of subtype B, C and D sequences. However later sequences were identified with a higher level of recombination and repeated detection of URFs (Engelbrecht et al., 2011).

1.3 The Structural and functional importance of HIV-1 tat

1.3.1 HIV-1 *tat* gene and Tat protein

HIV-1 Tat is an early and late regulatory protein essential for HIV-1 replication. The protein is cationic in nature and exists in two forms made of 86 and 101 amino acids respectively. The dualistic form of Tat is generated by the translation of differentially spliced viral transcripts. However the exon 2 form of Tat appears to dominate early HIV-infection while the exon 1 form is produced after rev protein expression. Truncated forms of the protein appear to be viral strain dependant and have been shown to induce the same biological effects as the complete protein sequence (Buscemi et al., 2007; Amendt et al.,1994; Jeang et al., 1999). The truncated form (86 AA) is the most studied form of HIV-1 Tat and is predominantly found in HIV-1 subtype B, while the 101 amino acid form of Tat can be found in subtype C and most other HIV-1 strains.

The Tat protein is between 14 and 16kDa in size and consists of 6 domains. The first 5 domains are found within exon 1 and the 6th domain is found within exon 2 as illustrated in **Figure 1.6**.

The N-terminal is the first domain (1-20 AA) and is said to tolerate mutations well. Domain 2 (21-40 AA) regulates Tat function, is highly conserved and contains 7 cysteine's. Mutations in this area render the protein non-functional. Domain 3 (41-48 AA) is called the core domain and is essential for the transactivation activity of Tat. Domain 4 (49-57 AA) is the basic domain and is the most studied of all the domains and contains a conserved RKKRRQRRR motif. Domain 5 is called the glutamine rich domain (58-72 AA) while the remaining residues complete the C-terminal (73-101 AA). The basic and C-terminal domains collectively confer Tat with TAR RNA binding capabilities which is necessary for both nuclear localization and the assimilation of Tat into cells (Buscemi et al., 2007; Jeang, 1996; Jeang et al., 1999; Pugliese et al., 2005). Tat is also a mediator of HIV-1 neuro-pathogenesis and does so with the use of neurotoxic and chemotactic domains encoded within the first exon (Cowley et al., 2010).

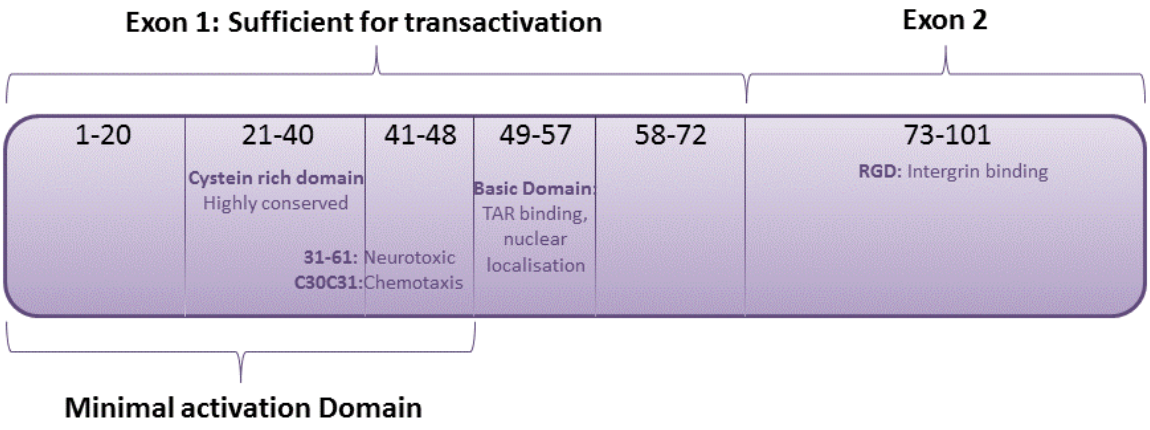

Figure 1.6: The domains of HIV-1 *tat*. The figure illustrates *tat* exon 1; *tat* exon 2 and the domains each are comprised of. The figure also highlights areas needed for integrin binding and nuclear localization and how many domains can sustain minimal or sufficient transactivation. (Adapted from: Kiebala, 2010).

1.3.2 Activation of HIV LTR by Tat

Early HIV LTR reporter gene studies showed that in the presence of Tat, transactivation activity was a hundred times higher than when Tat was absent (Cann et al., 1985). Tat has since been extensively studied and the mechanism of transactivation is well defined. Tat LTR activation was shown to be a RNA-dependent process; however other mechanisms of LTR activation include the use of transcription factors such as NF-κB, Sp-1 and NF-IL6. In the presence of Tat the LTR will produce long viral transcripts that enhance viral activation and in the absence of Tat, short viral transcripts will be produced. The activation of the LTR commences with the recruitment of the RNA Polymerase II complex. This complex consists of mediators and a carboxyl-terminal domain (CTD) that blocks the activation of the LTR. Tat activates cyclin-dependent kinase 7 (CDK7), a component of the TFIIH complex, resulting in the phosphorylation of the CTD and the removal of the mediators. Transcription of the tat activation region (TAR) is initiated, resulting in the TAR RNA to form a stem loop structure which facilitates the attachment of the RNA binding domain. The process of Tat and TAR binding is driven by a kinase complex, consisting of trans-kinase 9 (CDK9) and cyclin T1, which forms a zinc-dependent complex with Tat. CDK9 phosphorylates the CTD which relocates the TAR and allows the RNA polymerase to transcribe the HIV-1 genome (Karn, 1999)

1.3.3 Production of Tat

Since Tat is pivotal for HIV-1 infection it is expressed during the early stages of infection at low levels without the HIV genome being transcribed. Tat is also released from un-ruptured HIV-1 infected cells into the cell microenvironment (Arya et al., 1985; Westerndorp et al., 1995; Xiao et al., 2000). It has been suggested that Tat is not secreted as a single peptide; it does not make use of the endoplasmic reticulum for protein secretion but a pathway similar to what Interleukin-1 beta (IL-1β) and Basic fibroblast growth factor (bFGF). It has been reported that in HIV-1 infected cultured supernatants approximately, 0.1 to 1 ng/ml of Tat can be produced (Westerndorp et al., 1995). In addition the levels of intact functional Tat detected in patient serum appear to be proportional to viral titers found in the circulation (Finzi & Siliciano, 1998). It has therefore been speculated that Tat production is directly proportional to the progression of HIV-1. In addition, other studies have shown that

extracellular Tat is able to up-regulate HIV-1 replication while anti-Tat antibodies are said to do the reverse (Zauli et al., 1996). These findings therefore give credence to the theory that extracellular Tat may play a role in the progression of AIDS.

1.3.4 Cellular Uptake of HIV-1 Tat

Studies have shown that HIV-1 Tat is easily assimilated by various cells and is able to bind to a number of cell receptors such as CD26 CXCR4; heparin sulfate proteoglycans and low-density lipoprotein. These cell receptors often undergo endocytosis in T-cells using one of four endocytosis pathways: clathrin-mediated endocytosis, caveolae, micropinocytosis or phagocytosis (Gutheil et al., 1994; Xiao et al., 2000; Tyagi et al., 2001). T-cells are reported to use the clathrin-dependent pathway, while Tat utilizes the alveolar endocytosis pathway and clathrin-dependent pathways. This illustrates Tat's ability to utilize different cell receptors for cell entry (Fittipaldi et al., 2003; Rayne et al., 2004; Vendeville et al., 2004).

1.3.5 HIV-1 Tat mediated gene expression

The role of Tat is not limited to HIV-1 replication. It also plays a role in gene expression and is able to induce the expression of a number of cytokines. Tat is able to activate the TNF-α promoter which in-turn initiates the production of TNF-α and IL-10. TNF-α is responsible for the induction of apoptosis and other cellular responses while IL-10 inhibits Th1 cytokines thereby favoring viral replication (Sawaya et al., 1998; Badou et al., 2000). Furthermore, Tat is able to induce the expression of IL-6 which has been reported to exacerbate HIV-1 associated infections in infected patients (Kelly et al., 1998). A number of studies have also implicated Tat in the induction and suppression of apoptosis and have shown Tat to activate cell survival genes such as Bcl-2 while simultaneously inhibiting the expression of tumor suppressor gene p53 (Li et al., 1995; Zauli et al., 1996).

1.3.6 Apoptosis induction by HIV-1 Tat

Apoptosis is a controlled series of cell deaths. In normal cells, apoptosis facilitates the sculpturing of tissues and organs and is a critical aspect of organism development (Bouillet & Strasser, 2002; Meier et al.2000). It ensures the maintenance of cell populations as cells are often produced in excess. It is therefore

critical for homeostasis. In most instances, apoptosis is advantageous to an organisms life cycle, however, abnormalities in apoptosis regulation results in the development of various diseases including cancer (O'Connor et al., 2000; Romani et al., 2010; Strasser et al., 2000). Apoptosis is controlled by a number of cell signals, which induce one of two apoptotic pathways. The extrinsic pathway is triggered by external stimuli and ligands, which include toxins, cytokines and death receptors such as Fas/CD95. The intrinsic pathway is regulated by mitochondria and the release of apoptotic-inducing molecules such as cytochrome *c* (Bouillet & Strasser, 2002; Romani et al., 2010). These cell signals can either actively induce apoptosis or inhibit the apoptotic process (Denault & Salvesen, 2002). Studies have identified four groups of compounds or cell signals which affect the apoptotic process and include caspases, adaptor proteins, members of the tumor necrosis factor receptor (TNF-R) super family and members of the Bcl-2 family (O'Connor et al., 2000; Strasser et al., 2000).

Tat induces the intrinsic pathway in various human cells by activating various apoptotic proteins. Tat activates caspase 3, caspase 8 and tumour necrosis factor-related apoptosis-induced ligand (TRAIL), in human monocytes. Tat also induces apoptosis by activating TNF-α and Fas ligands (FasL/CD95) in astrocytes. The latter is also responsible for the depletion of CD4 B cells. Furthermore, Tat also associates with microtubules and induces mitochondrial apoptosis. Induction of microtubule stabilization by Tat is potentially pro-apoptotic as cells require constant polymerization and depolymerization of microtubules. The alteration of microtubules acts as a strong inducer of the intrinsic apoptosis pathway (Romani et al., 2010). Many human neurodegenerative conditions are accompanied by a reorganization of the neuronal cytoskeleton and as tat interacts with micro tubular networks this could possibly explain the occurrence of HIV-1 neuropathogenesis. Recruitment of various apoptosis-inducing proteins by Tat has suggested the ability of the transactivation protein to utilize numerous pathways and cells for the induction of apoptosis. (Ashkenazi, 2002; O'Connor et al., 2000; Romani et al., 2010).

1.3.7 Apoptosis suppression by HIV-1 Tat

Inhibition of apoptosis by Tat seems to be in conflict with its role as an apoptosis inducer. However, Tat has been shown to up-regulate B-cell lymphoma 2 (Bcl-2), an

anti-apoptotic protein in monocytes. Extracellular Tat can be absorbed by peripheral blood monocytes and induce BcL-2 to inhibit apoptosis by TRAIL. Tat also associates with B-cell lymphoma-extra large (Bcl-xl), another member of the Bcl2-family and forms a compound capable of inhibiting apoptosis in neurons.The anti-apoptotic properties of TAR could possibly explain how infected cells could inhibit apoptosis while extracellular Tat induces apoptosis. The anti-apoptotic activities of Tat could therefore cause the formation of HIV-1 latency in human cells, although Tat apoptosis in the CNS is complex (Romani et al., 2010).

1.3.8 Mutations in *tat* exon 1

Clade-specific differences in transmission and spread of the various subtypes have been documented. However, it is unclear if differences in biological properties may influence disease progression (Mishra et al., 2008). Studies conducted on subtype B and subtype C analyzed sequences on a genetic and molecular level and provided evidence that a point mutation (C30S31) of subtype C decreases the chemotactic functions of Tat. The mutation is not geographically limited and has been identified in a number of different subtype C strains as well as other non-subtype B sequences (Mishra et al., 2007). In addition, subtype C Tat differs from subtype B Tat due to a C31S mutation (Campbell et al., 2005; Contreras et al., 2005; Desfosses et al., 2005). It has been reported that Tat is a chemo attractant for macrophages, monocytes and dendritic cells and induces chemotaxis via a CCR2-dependent mechanism that is dependent upon the integrity of the dicysteine motif of Tat. The cysteine-rich domain of Tat is reported to contain positions of amino acid similarity with β-chemokine critical for chemokine receptor binding and signal transduction. The C30S31 mutation terminates the ability of Tat to act as a chemo attractant by causing inhibition of the CCR2-dependant mechanism. Furthermore, Tat also exerts a neurotoxic effect on hippocampal neurons. Tat and neurocognitive impairment has been shrouded in controversy, but there is now evidence that Tat mediates neurotoxicity via its cysteine and basic domains. Yet, this is not the only route of inducing neurotoxicity. It can also exert a neurotoxic effect on neurons by uncoupling Ca^{2+} permeable N-methyl-D-aspartate (NMDA) receptors from Zn^{2+} inhibitors, thereby causing NMDA mediated death (Rao et al., 2008; Campbell & Loret, 2009).

In addition, Tat exon 1 contains a number of adjacent point mutations which may influence the transactivation and phosphorylation of the virus, summarized in **Figure 1.8** and **Table 1.2** (Cowley et al., 2010). According to Jeang, 1996 the majority of mutations are concentrated within amino acids 1-58 However, analysis of mutations revealed that the first 21 amino acids are fairly tolerant of mutations while mutations in amino acids 22-40 was problematic and often detrimental for trans-activation. In contrast mutations in amino acids 49-57 did not significantly alter Tat function (Jeang, 1996).

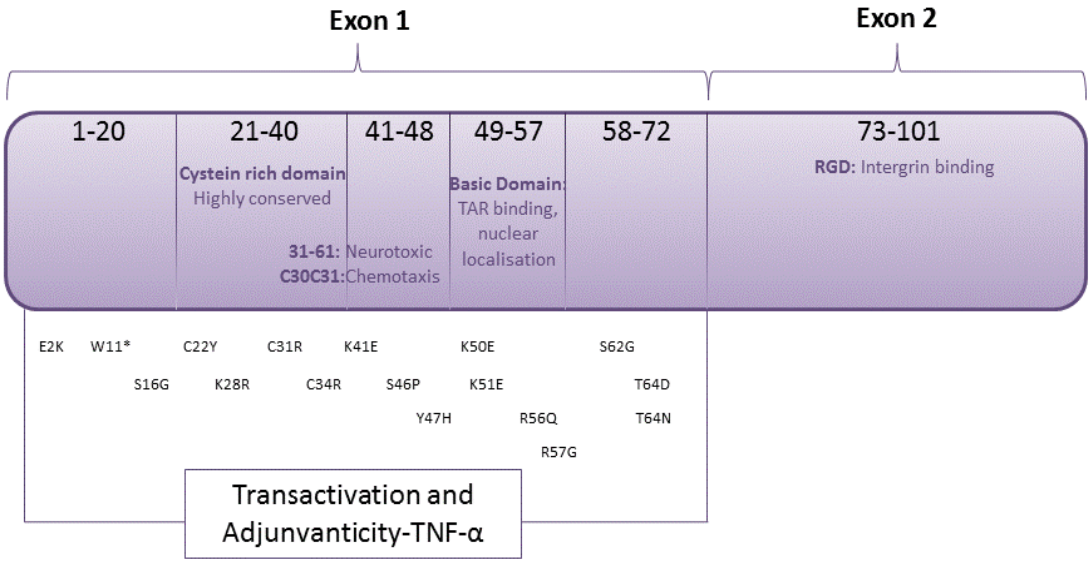

Figure 1.8: Summary of tat mutations identified in HIV-1 clade B sequences. A total of 17 mutations were identified from the literature. The majority of these mutations have been associated with transactivation activity. (Source: Adaption of Kiebala, 2010 and Cowley et al., 2010).

Table 1.2 Significance of amino acid substitutions Tat exon 1 (Cowley et al., 2010).

Amino Acid substitution	**Importance**
E2K	E2 provides intermolecular hydrogen bond with K51 and R53
W11*	Stop codon, truncated Tat protein
S16G	S16 site of phosphorylation by CDK2, mutants associated with reduced transactivation
C22Y	C22 required for transactivation
K28R	K28 site of Tat acetylation by p300/PCAF, potentially decreased transactivation
C31R	C31 required for transactivation
C34R	C34 required for transactivation
K41E	K41 mutation reduces transactivation
S46P	S46 site of phosphorylation by CDK2, associated with reduced transactivation
Y47H	Y47H associated with reduced transactivation
K50E	K50 site of Tat acetylation by p300/PCAF, potentially decreased transactivation
K51E	K51 site of acetylation by p300/PCAF, potentially decreased transactivation
R56Q	Reduced TAR interaction and transactivation
R57G	Reduced TAR interaction and transactivation
S62G	S62 site of phosphorylation by PKR, associated with decreased TAR interaction
T64N	T64 site of phosphorylation by PKR, associated with decreased TAR binding
T64D	T64 site of phosphorylation by PKR, associated with decreased TAR

1.4 Neurocognitive complications of HIV-1

1.4.1 HIV associated neurocognitive disease (HAND)

According to Antinori et al. (2007), HAND can be divided into three categories: asymptomatic neurocognitive impairment (ANI), minor neurocognitive disorder (MND) and HIV-1 associated dementia (HAD). ANI and MND are the most frequently occurring forms of HAND and affect 30%-60% of HIV-1 infected individuals (Grant, 2008). ANI is defined by subtle cognitive impairment and occurs during the early stages of HIV-1 infection and can last for up to 10 years. MND is similar to ANI however individuals experience mild impairments such as motor function and behavioral abnormalities which may affect daily activities.

HAD occurs during the final stages of HIV-1 infection and is characterized by extreme difficulties in attention, concentration, speed of processing information, mental flexibility, verbal learning and memory, as well as motor speed (Gupta et al., 2007; Grant, 2008).

HAND has slowly become more prevalent in HIV-1 infected individuals and is a growing concern for the global HIV-1 population. A number of HAND cases have been reported in America and recently in South Africa as well. HAND has been described as a culmination of genetic and cellular factors such as neurotoxic proteins (gp41, Tat) and cytokines (IL-6 and TNF-α) which results in a varied spectrum of clinical manifestations (Ranga et al., 2004).

1.4.2 Prevalence of HAND

It is estimated that 15 000 new cases of HAND including ANI, MND and HAD are reported annually (Odiase et al., 2006). The prevalence of neurocognitive impairment has not been clearly defined as some studies have reported a prevalence of 5% while others have reported a prevalence of 20%. A study conducted in 2007 found that 60.5% of study participants experienced mild to moderate cognitive deficits associated with memory or learning (Gupta et al., 2007). Similarly, Grant (2008) reported that 50% of HIV-1 treatment-naive patients in their cohort as having a form of cognitive, behavioral or motor impairment.

1.4.3 HAART and HAND

It has been reported that although Highly Active Anti-retroviral Therapy (HAART) effectively dampens the symptoms of HAND it is unable to prevent the eventual onset of the neurological disorder (Tardieu, 1999). The CNS acts as a reservoir for the virus replication even during maximal treatment with HAART. It is therefore not surprising that HIV-1 associated neurocognitive disorders persist (Boissé et al., 2008; Jevtović et al., 2009). However HAND is not fully understood and various mechanisms have been implicated in the progression and severity of the disease (Mishra et al., 2008).

HAART, a combination therapy, consisting of three classes of drugs: protease inhibitors (PI), nucleoside reverse transcriptase inhibitors (NRTI), and non-nucleoside reverse transcriptase inhibitor (NNRTI) is said to lower the symptoms of HAND although its effects are disputable for a number of reasons (Parsons et al., 2006; Sacktor et al., 2006). Due the protective nature of the blood brain barrier (BBB) the CNS is not easily accessible to invasive chemicals such as antiretroviral drugs (Lin et al., 1996; Brew, 2001). Due to the highly protein-bound nature of PI, very little or none of them are translocated across the BBB. NRTIs and NNRTIs have displayed moderate penetration of the BBB but only at hematologic intolerable doses (Brouwers et al., 1997; Gisslén et al., 1997; Brew, 2001).

1.4.4 HAND and HIV-1 subtypes

A number of studies have suggested that the severity and prevalence of HIV-1 associated cognitive impairment are clade specific (Sacktor et al., 2006; Mishra et al., 2008). It has been proposed that cognitive impairment especially HAD is most prevalent in subtype B due to the presence of the C30C31 mutation in the Tat motif. However, one would expect more HAND in sub Saharan Africa especially given the frequency of opportunistic infections such as tuberculosis, malaria and poverty to complicate existing HIV-1 infection (Sacktor et al., 2006). In addition subtype C is also responsible for more than 50% of the world's HIV-1 infections particularly in developing countries such as India, Africa and China (Liner et al., 2007; Valcour, et al., 2007). Studies conducted in India proposed that subtype C was associated with reduced levels of cognitive impairment than subtype B as none of the study

participants presented with any symptoms of HAD (Chandra et al., 2006; Gupta et al., 2007). However a study conducted in Cape Town, South Africa found subtype C HIV- 1 associated cognitive impairment to be prevalent with HAD manifesting in 23.5% of the study participants (Joska et al., 2010). This reinforces Sacktor et al.'s, (2005) claim that HAD may become the leading cause of dementia in young adults.

1.4.5 Role of HIV-1 Tat in HAND

A growing amount of evidence has cemented the link between HAND and HIV-1 Tat, which has been described as a toxic protein capable of inducing apoptosis via a multitude of pathways. Tat is able to easily enter various cell types including T and B cells, as well as cells of the CNS (Pu et al., 2003). The release of Tat by infected cells results in the increase of TNF-α, induction of oxidative stress in the cellular environment and the migration and adhesion of endothelial cells activated with T helper-1 type inflammatory cytokines. It also results in the activation of caspases and ultimately apoptosis of neurons. The induction of oxidative stress is most likely a product of excess Ca^{2+} in neurons and the production of reactive oxygen species (ROS) by mitochondrial calcium uptake. Oxidative stress has been speculated to be a major cause of neurodegenerative diseases and has been detected in both the brain and cerebrospinal fluid (CSF) of HAND patients (Pu et al., 2003; Buscemi et al., 2007; Romani et al., 2010).

AIMS AND OBJECTIVES OF THE STUDY

The aims of this project was to subtype HIV-1 in the Cape Town cohort being investigated for HAND with a special focus on characterizing the HIV-1 *tat* gene and Tat protein *in vitro*.

The specific objectives were to:

1. Subtype HIV-1 in the patient cohort.
2. Characterize the HIV-1 *tat* gene in the HAND cohort and identify mutations in the C30C31 position.
3. Construct recombinant *tat* clones and investigate the apoptotic ability of the clones in cell culture.

To achieve the first objective, HIV-1 nucleic acids were isolated, amplified, sequenced and subtyped using phylogenetic analysis as described in **Chapter 2**.

The Tat sequences obtained from the HAND cohort were then further analyzed and compared with different subtype B and subtype C datasets (**Chapter 3**).

Specific mutations observed in the C30C31 position were then used to construct recombinant clones in an inducible CMV expression vector. After transfection of 293T cells with the clones, the subsequent expression of Tat protein was assessed with flow cytometry to determine the apoptotic ability of each mutation (**Chapter 4**).

CHAPTER TWO

HIV-1 genetic diversity: Molecular characterization of partial *gag*, *pol*, *accessory* and *gp41* fragments

2.1. Introduction

In this chapter the subtyping of the HIV-1 strains in the cohort are described. Briefly, after the isolation of viral RNA and proviral DNA, the nucleic acids were used as templates for the amplification of different genomic regions of the partial *gag*; *pol*; accessory and *gp41* genes. Pre-nested (first round) RT-PCR assays were used for the reverse transcription of RNA and amplification of cDNA and PCR were used for first round amplification of DNA. After nested (second round) PCR, the amplified gene fragments were visualized with agarose gel electrophoresis and directly sequenced. Sequences were verified and used in the phylogenetic analysis of the different HIV-1 genes. An overview of the different techniques used, is indicated in **Figure 2.1.**

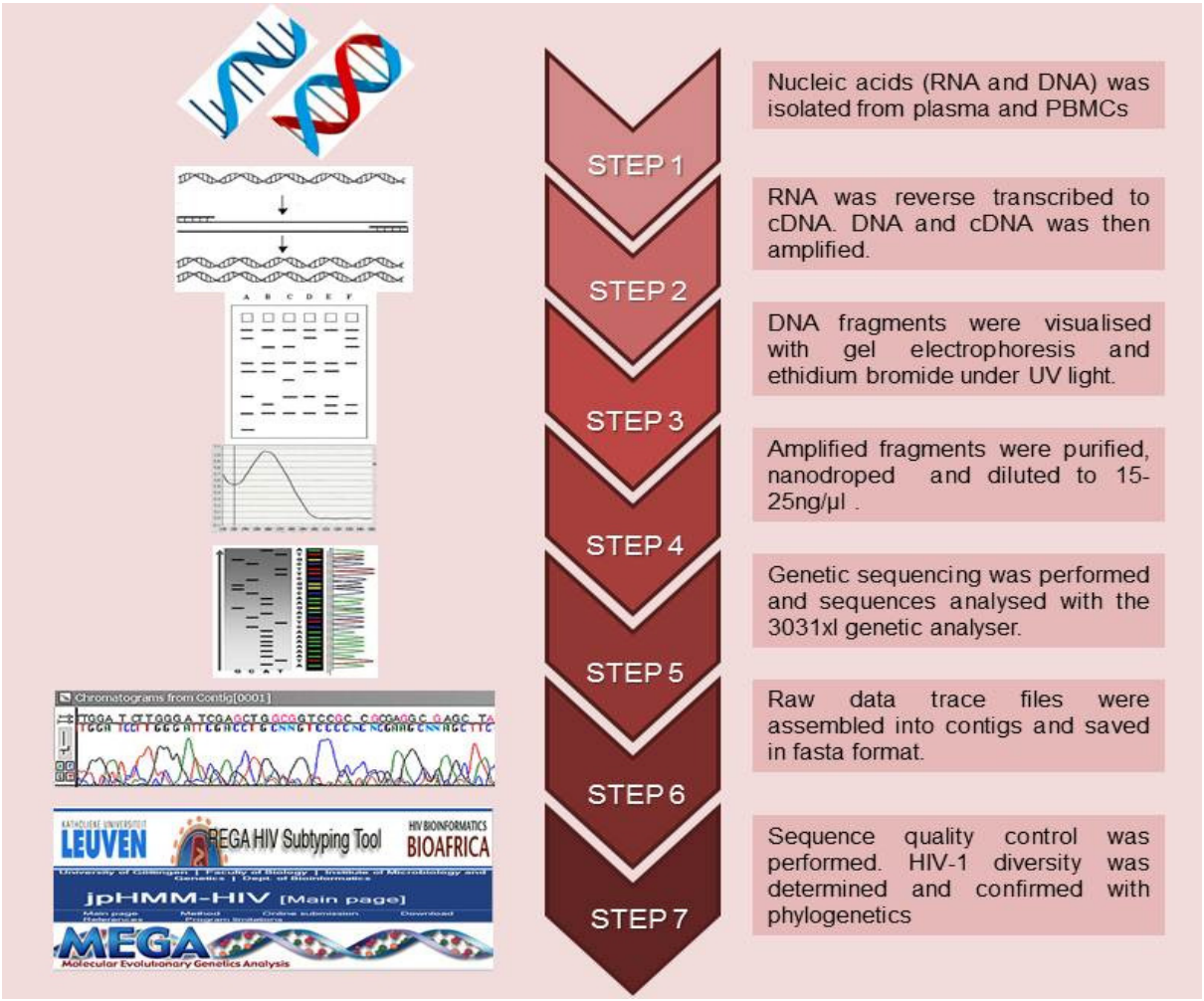

Figure 2.1: Overview of steps used to determine HIV-1 subtypes and genetic diversity.

2.2. Materials and Methods

2.2.1 Reagents and equipment

All the reagents, equipment, and software applications that were used during the course of this study are listed in **Tables 2.1 - 2.3**. Registered trade mark items and trade mark products are indicated by the symbols ® and TM respectively.

Table 2.1. List of chemicals and kits used in the study.

Chemical or Commercial products and kits used	Supplying Company
QIAamp Viral RNA Mini Kit	QIAGEN, Dusseldorf, Germany
QIAamp DNA Blood Mini Kit	QIAGEN, Dusseldorf, Germany
Access RT-PCR System	Promega, Madison, WI, USA
GoTaq® DNA Polymerase	Promega, Madison, WI, USA
dNTP´s	Promega, Madison, WI, USA
Nuclease free water	Promega, Madison, WI, USA
Agarose	Whitehead Scientific (Pty) Ltd.
Ethidium Bromide	Promega, Madison, WI, USA
6x Blue Orange Loading Dye	Promega, Madison, WI, USA
QIAquick PCR Purification Kit	QIAGEN, Dusseldorf, Germany
BigDye™ Terminator sequence Kit	Applied BioSystems, CA, USA
5x Sequencing Buffer	Applied BioSystems, CA, USA
BigDye XTerminator Purification Kit	Applied BioSystems, CA, USA
Exonuclease (Exo I)	Affymetrix, OH, USA
Shrimp alkaline phosphate	Affymetrix, OH, USA
ExoSAPIT	Affymetrix, OH, USA
1Kb DNA Ladder	Promega, Madison, WI, USA

Table 2.2. Equipment used to perform sample analysis.

Piece of Equipment	Supplying Company
QIAcube nucleic acid isolation system	QIAGEN, Dusseldorf, Germany
GeneAmp PCR System 9700 thermal cycler	Applied BioSystems, CA, USA
Hoeffer EPS 2 A 200, Power Pack	Pharmacia Biotechnologies, CA, USA
Syngene™ GeneGenius Computer System	Synoptics Ltd., Cambridge, UK
ABI 3130xl automated Genetic analyser	Applied BioSystems, CA, USA
Nanodrop™ ND 1000	Nanodrop Technologies Inc. Delaware, USA
Techne® Gene E Thermal Cycler	Techne Ltd. Cambridge, UK

Table 2.3. Software programs and online tools that were used in sequence analysis.

Software package	References and/or licensed companies
Sequencher v 4.8	Gene Codes Corporation, Ann Arbor, MI, USA
ClustalW v 2.1	Thompson© *et al,* 1997
LANL QC	Los Alamos National Security, USA
REGA v 3.0 HIV subtyping tool	de Oliveira *et al.,* 2005
jpHMM	Spang *et al.,* 2002
SQUEAL	Pond *et al.,* 2009
Stanford	Rhee *et al.,* 2003
MEGA 5.1	Tamura *et al.,* (2011)
Geneious	Biomatters Ltd, Auckland, NZ
CPR	Gifford *et al.,* 2009

2.2.2 Patient samples

Patients (n = 64) were enrolled from 3 clinics in the Cape Town area and were assigned numbers to ensure confidentiality e.g. (JO198). Stored samples were kindly provided by Prof. John Joska (UCT) and consisted of 2ml Peripheral Blood Mononuclear Cells (PBMCs) and 6ml of plasma for each patient. Samples were stored at -20^{o}C until used for nucleic acid extractions.

2.2.3 Nucleic acid extractions

Viral RNA extractions:

RNA was extracted from 140µl of plasma using the QIAamp mini viral kit (Qiagen GmbH, Hilden, Germany) using the QIAGene QIAcube automated extraction system. The kit protocol was followed and RNA was eluted in 60µl of low salt buffer AVE.

The Qiagen RNA extraction makes use of a silica-gel membrane principle developed by Vogelstein and Gillespie (1979). Lysing buffer is added to patient plasma to degrade the viral capsid and to release nucleic acids, which are then bound to a silica membrane within a QIAamp spin column during centrifugation. Residual contaminants were removed from the silica membrane with wash steps to ensure the purity of the sample. Salt and pH conditions help to retain nucleic acids within the silica membrane and ensure that they could only be eluted with water or low salt buffers. Isolated RNA was used for the molecular characterization of *gag* p24 and *pol* gene fragments.

Proviral DNA extractions:

DNA was extracted from 200µl PBMCs using the QIAamp DNA blood mini kit (Qiagen GmbH, Hilden, Germany) using the QIAGene QIAcube automated extraction system. The kit protocol was followed and DNA was eluted in 100µl of low salt buffer AE. Isolated DNA was used for the molecular characterization of the accessory gene and *gp41* gene fragments.

2.2.4 Polymerase chain reactions

All samples were amplified for the four gene regions: *gag, pol, accessory genes* and *gp41* as indicated in **Figure 2.2**.

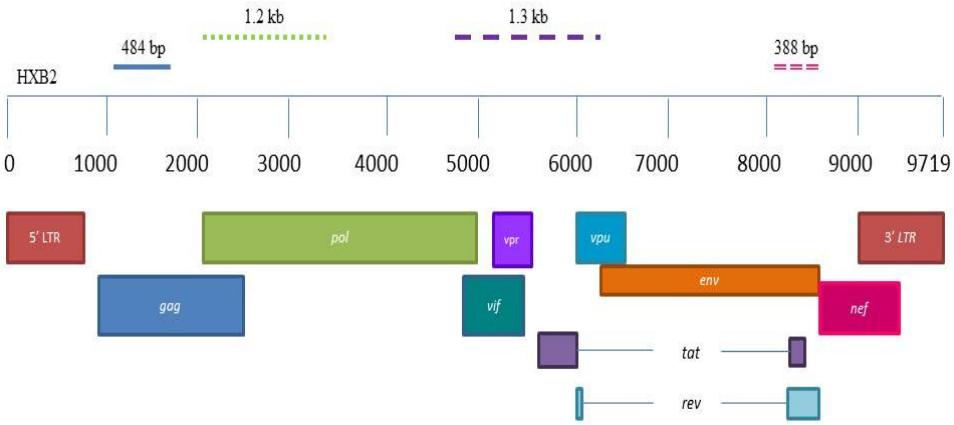

Figure 2.2 Amplified gene fragments. Four gene fragments were amplified for patient samples partial *gag* (484bp), *pol* (1.2kb) *accessory* (1.3kb) and *gp41 (*388bp) genes. Region and size of amplified fragments are indicated by lines on HXB2 scale.

2.2.4.1 Partial *Gag* RT-PCR

A total of 5µl of RNA was used for pre-nested (first round) *gag* p24 reverse transcriptase polymerase chain reaction (RT-PCR). The RNA was converted to cDNA by using the Access RT-PCR system (Promega, USA) and then used as the template in a one-step PCR reaction.

The primer pairs used for pre-nested and nested *gag* amplification are described in **Table 2.4** and were used at a final concentration of 40 pmol. Final concentrations of the remaining reagents were as follows: 0.2mM dNTP's, 1mM $MgSO_4$, 1.1U AMV/ *Tfl,* 1X reaction buffer and 1.1U AMV reverse transcriptase. Nuclease-free water was added to a final volume of 50µl. A second nested PCR was performed, using 3µl of the pre-nested PCR product as template. The final concentrations of reagents for nested PCR reaction were as follows 0.2mM dNTPs, 1.5mM $MgCl_2$, X1 GoTaq Flexi buffer, 0.05U Taq DNA polymerase, water was added to a final volume of 50µl per reaction. The Gene Amp® 9700 PCR system (Applied Biosystems, USA) was used for the amplification of *gag* fragments with cycling parameters described in **Table 2.5**

Table 2.4. Primers used in the amplification of the partial HIV-1 gag gene products.

	Primer	Oligonucleotide sequence	F/R	HXB2 Location	T_M (℃)
cDNA and first round	p24-7	CCCTGRCATGCTGTCATCA	R	1826←1844	55.2
	p24-1	AGYCAAAATTAYCCYATAGT	F	1174→1193	56.3
Nested PCR	p24-2	AGRACYTTRAAYGCATGGGT	F	1237→1256	50.0
	p24-6	TGTGWAGCTTGYTCRGCTC	R	1703←1721	50.0

Table 2.5 Cycling parameters of *gag* p24 PCR

Cycling conditions for the PreNested *gag* p24 PCR assay			
Step	**Temperature**	**Duration**	**Cycles**
Reverse Transcription	48℃	45 min	x1
Initial Denature Step	94℃	2 min	x1
Denature	94℃	20 sec	x40
Anneal	45℃	30 sec	
Extend	68℃	90 sec	
Final Extension	68℃	10 min	x1
Cycling conditions for the Nested *gag* p24 PCR assay			
Step	**Temperature**	**Duration**	**Cycles**
Initial Denature Step	94℃	2 min	x1
Denature	94℃	20 sec	x40
Anneal	50℃	30 sec	
Extend	68℃	60 sec	
Final Extension	68℃	10 min	x1

sec – seconds; min – minutes; ℃ – Degrees Celsius

2.2.4.2 Partial Pol RT-PCR

The PCR procedures described in Section 2.3.1 was repeated for the *pol* gene with 5µl of RNA being used for the pre-nested *pol* RT-PCR reaction and 3µl of this product being used as template for the nested PCR. Concentrations of reagents and primer pairs were the same as described before. The primer pairs and cycling parameters are described in **Tables 2.6 and 2.7** respectively.

Table 2.6: HIV-1 *pol* gene amplification primers

	Primer	Oligonucleotide sequence	F/R	HXB2 Location	T_M (℃)
cDNA and first round	PR-5′ prot -1	TAATTTTTTAGGGAAGATCTGGCCTTCC	F	2082→2109	57.0
	RT-MJ4	CTGTTAGTGCTTTGGTTCCTCT	R	3399←3420	55.0
Nested PCR	PR-5′ prot-2	TCAGAGCAGACCAGAGCCAACAGCCCCA	F	2136→2163	68.0
	RT-NE135	CCTACTAACTTCTGTATGTCATTGACAGTCCAGCT	R	3300←3334	62.0

Table 2.7: Cycling parameters of *pol* PCR

Cycling conditions for the PreNested RT-*pol* PCR assay			
Step	Temperature	Duration	Cycles
Reverse Transcription	65℃	30 sec	x1
Initial Denature Step	94℃	2 min	x1
Denature	94℃	20 sec	x40
Anneal	56℃	30 sec	
Extend	68℃	90 sec	
Final Extension	68℃	5 min	x1
Cycling conditions for the Nested RT-*pol* PCR assay			
Step	Temperature	Duration	Cycles
Initial Denature Step	94℃	2 min	x1
Denature	94℃	30 sec	x40
Anneal	65℃	30 sec	
Extend	68℃	60 sec	
Final Extension	68℃	7 min	x1

2.2.4.3 Accessory genes PCR

The PCR amplified a sub-genomic region encompassing five genes namely *vpr*, *vif*, *tat*, *rev* and *vpu* using 5µl of proviral DNA for the pre-nested PCR and 3µl of this product for the nested PCR. Both pre-nested and nested PCR's were performed using the Promega GoTaq Flexi ® kit (Promega, USA) and contained 0.2mM dNTP's, 1.5mM $MgCl_2$, X1 GoTaq Flexi buffer, 1.1U Taq DNA polymerase and nuclease free water in a final volume of 50µl. Primers for both pre-nested and nested PCR's are

summarized in **Table 2.8** and were used at a final concentration of 40pmol. Cycling parameters for both PCRs are described in **Table 2.9**.

Table 2.8: Summary of primers used for the accessory gene amplification

	Primer	**Oligonucleotide sequence**	**F/R**	**HXB2 Location**	**T_M (°C)**
cDNA and first round	FGF46	5' -GCA TTC CCT ACA ATC CCC AAA G-3'	F	4648-4669	56.1
	ES8X	5' -CAC TTC TCC AAT TGT CCC TCA-3'	R	7668-7648	55.7
Nested PCR	N817	5' -TCT GGA AAG GTG AAG GGG CAG T-3'	F	4954-4975	60.7
	N1156	5' -TCA TTG CCA CTG TCT TCT GCT CT-3'	R	6206-6228	58.5

Table 2.9 Cycling parameters of accessory gene amplification

Cycling conditions for the PreNested RT-*pol* PCR assay			
Step	Temperature	Duration	Cycles
Initial Denature Step	94°C	2 min	x1
Denature	94°C	30 sec	x40
Anneal	55°C	30 sec	
Extend	68°C	3 min	
Final Extension	68°C	10 min	x1
Cycling conditions for the Nested RT-*pol* PCR assay			
Step	Temperature	Duration	Cycles
Initial Denature Step	94°C	2 min	x1
Denature	94°C	30 sec	x40
Anneal	55°C	30 sec	
Extend	68°C	90 sec	
Final Extension	68°C	7 min	x1

2.2.4.4 Partial *env gp41* PCR

Partial env *gp41,* including the *tat exon* 2 and *rev exon 2* fragments, were amplified using primers described in **Table 2.10** at a final concentration of 40pmol. Concentrations of pre-nested and nested PCR reactions were as follows: 0.2mM dNTP's, 1.5mM $MgCl_2$, X1 GoTaq Flexi buffer, 1.1U Taq DNA polymerase and Nuclease free in a final volume of 50µl. All PCR's were performed on the GeneAmp® 9700 PCR system (Applied Biosystems, USA) using 5µl and 3µl of DNA for pre-nested and nested PCR's respectively. Cycle parameters have been summarized in **Table: 2.11**.

Table 2.10. Summary of primers used for the amplification of *gp41* gene fragments.

	Primer	**Oligonucleotide sequence**	**F/R**	**HXB2 Location**	**T_M (℃)**
cDNA and first round	N1169	5'-AAT ATT CAT AAT GAT AGT AGG AGG-3'	R	8273-8296	47.4
	N847	5'-TTA TTG CAA AGC TGC TTC AAA GCC CTG TC-3'	F	8767-8795	61.6
Nested PCR	Rev X2R	5'-TCCTATCTGTTCCTTCAGCTA-3'	R	8709-8689	60
	Rev X2F	5'-TGCTGTGCTCTCTATAGTRA-3'	F	8321-8340	55

Table 2.11. Cycling parameters of gp41 gene fragment amplification

Cycling conditions for the PreNested RT-*pol* PCR assay			
Step	Temperature	Duration	Cycles
Initial Denature Step	94℃	2 min	x1
Denature	94℃	30 sec	x40
Anneal	43℃	30 sec	
Extend	68℃	1 min	
Final Extension	68℃	7 min	x1
Cycling conditions for the Nested RT-*pol* PCR assay			
Step	Temperature	Duration	Cycles
Initial Denature Step	94℃	2 min	x1
Denature	94℃	30 sec	x40
Anneal	50℃	30 sec	
Extend	68℃	60 sec	
Final Extension	68℃	7 min	x1

2.2.5 Gel electrophoresis and purification of amplified fragments

Partial *gag*, *pol*, accessory gene fragments were visualized using 0.8% to 1% agarose gel electrophoresis. Tris-Acetate-EDTA (TAE) buffer (0.04 M Tris acetate, 0.001 M EDTA) was used to prepare agarose gels. DNA was stained with 5µl of ethidium bromide (0.5 µg per ml; Promega, USA) in a total volume of 50ml TAE buffer (1X). The marker used was a 1kb molecular weight DNA marker (Promega, USA) and also included was a negative and a positive control (sample previously positively amplified). DNA fragments were visualized with the UVItec Alliance Chemilluminscence and Fluorescence gel imaging system (Cambridge, UK). The expected fragment sizes for *gag p24*, *pol*, accessory gene and *gp41 gene* fragments were 484bp 1.2kb, 1.3kb and 388bp respectively. Visualized PCR products were purified using one of two methods, described below.

Purification of gel fragments using Exonuclease 1 (Exo) and Shrimp Alkaline Phosphate (SAP)

The first method utilized the enzymes Exonuclease 1 (Exo) and Shrimp Alkaline Phosphate (SAP) (USB Corporation, USA). A 10ul aliquot of the PCR products was placed into a 0.2-ml thin-walled PCR tube containing 0.5µl of SAP (1 U/µl) and 0.5µl Exonuclease (1 U/µl). These enzymes are able to degrade any substances which may inhibit the sequencing process. The SAP dephosphorylates the remaining deoxynucleotides and the exonuclease degrades primers and hydrolyses single stranded DNA. A Techne® Gene E Thermal cycler (Techne Ltd., UK) was used to incubate the reaction for 15 minutes at 37ºC. The enzymes were then inactivated at 80ºC for 15 minutes. Purified samples were stored at 4ºC before sequencing reactions were performed. The Exo / SAP method has been labeled as a very crude method of DNA purification and therefore a second method was also used for the larger fragments.

Purification of gel fragments using the SV gel and PCR clean-up system

The second method made use of the Wizard SV gel and PCR clean-up system (Promega, USA). Approximately 10µl of the PCR product was mixed with an equal volume of membrane binding solution and transferred to a Wizard SV mini-column. Remaining PCR product was stored at -20ºC if purification was insufficient. DNA was

bound to the spin column membrane by centrifugation and washed twice to remove any impurities that may still be present. DNA was eluted with 50µl of nuclease free water. DNA concentrations were determined using the Nanodrop™ ND-1000 system (Nanodrop Technologies Inc.; USA) using 1µl of DNA. DNA concentrations are determined by measuring the sample absorbance (A) at a wavelength of 260 nm. The system can also determine the purity of the sample by dividing the absorbance at 260 nm by the absorbance of the sample at 280 nm. A purity reading between 1.7 and 1.9 is indicative of pure DNA which is free of contaminants and proteins (Sambrook et al.,1989) (Ausubel *et al*, 2003).

2.2.6 Sequencing of the amplified fragments

Amplifiable *gag* and *pol* gene fragments were directly sequenced at a concentration of 50ng /µl while accessory gene and *gp41* gene fragments were sequenced at a concentration of 20ng/ul. Primer pairs used for the sequencing reactions are summarized in **Table 2.12** and were diluted with nuclease free water (Promega, USA) to a concentration of 5pmol/µl and used together with the BigDye™ Terminator v3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, USA). Each sequencing reaction consisted of 1.3µl of BigDye Terminator Enzyme Mix, 2.7µl of Half Dye (Bioline USA Inc., USA), 1µl of primer, 1µl of purified DNA and 4µl of nuclease free water (Promega, USA) to a final volume of 10µl. Cycling parameters are indicated in **Table 2.13.**

Table 2.12: Summary of primers used for the sequencing of fragments.

Primers and Cycling conditions used for the sequencing of *gag* p24 PCR products			
Primer	**Forward / Reverse**	**Annealing Temperature**	**Reference**
gag p6	Reverse	53°C	Swanson *et al.,* 2003
gag p2	Forward	53°C	Swanson *et al.,* 2003
Primers and Cycling conditions used for the sequencing of *pol* PCR products			
Primer	**Forward / Reverse**	**Annealing Temperature**	**Reference**
JA 217	Reverse	60°C	Plantier *et al.,* 2005
30 prot 2	Reverse	60°C	Plantier *et al.,* 2005
Pol 1D	Forward	60°C	A Loxton personal communication
AK 10	Forward	60°C	Plantier *et al.,* 2005
AK 11	Forward	60°C	Plantier *et al.,* 2005
pol 3	Forward	60°C	S Engelbrecht personal communication
NE 135	Reverse	60°C	Plantier *et al.,* 2005
Primers and Cycling conditions used for the sequencing of *accessory* PCR products			
Primer	**Forward / Reverse**	**Annealing Temperature**	**Reference**
Vpr R1	Reverse	55	S Engelbrecht personal communication
Vif 1F	Reverse	55	S Engelbrecht personal communication
N111	Forward	55	Dr V Prasad personal communication
N1156	Forward	55	Dr V Prasad personal communication
Primers and Cycling conditions used for the sequencing of *env gp41* PCR products			
Primer	**Forward / Reverse**	**Annealing Temperature**	**Reference**
Rev X2R	Reverse	50°C	S Engelbrecht personal communication
RevX2F	Forward	50°C	S Engelbrecht personal communication

Table 2.13 Sequencing cycle parameters

	Temperature	Time	Cycles
Denaturation	96 ºC	10 seconds	25 X
Annealing	ºC*	5 seconds	
Extension	60 ºC	4 minutes	

*Annealing temperatures for sequencing primer sets were indicated in Table 2.12.

Gag and *pol* fragments were sequenced at the Central Analytical Facility at the Stellenbosch University, while accessory gene and *gp41* gene fragments were sequenced at the Department of Pathology at the Division of Medical Virology. The latter sequences were purified using the BigDye XTerminator Purification Kit and analyzed on an ABI 3130XL genetic analyser (Applied Biosystems, USA). This is a polymer-based capillary electrophoresis sequencer connected to a computer which uses sequencing analysis software (Applied Biosystems, USA) and captures raw

data trace files. Raw data trace files were retrieved from the Central Analytical Facility and the 3130XL ABI Prism genetic analyser and imported into Sequencher 4.8 (Gene Codes Corporation, USA). Low quality sequence ends were removed from overlapping nucleotide sequences to improve the quality of the sequences before being assembled into a single contig file. Assembled contig files were screened for ambiguities, edited and exported and saved in fasta format as a text file (.txt).

2.2.7 Sequence quality control

The online Quality control tool was used to examine the sequence sets for common problems: http://www.hiv.lanl.gov/content/sequence/QC/index.html Results from the tool include: subtype (from RIP), similarity (from HIV BLAST), phylogenetic (from Neighbor TreeMaker), stop codons and frameshifts (from GeneCutter) and hypermutation (from HyperMut).

2.2.8 HIV-1 characterization using online subtyping tools

2.2.8.1 Rega 3.0 subtyping analysis

REGA (http://bioafrica.mrc.ac.za:8080/rega-genotype-3.0.2/hiv/typingtool#/) is an online subtyping tool which utilizes phylogenetic analysis and bootscanning methods to screen sequences for recombination and subtype specific gene fragments. The genetic subtyping tool REGA 3.0 is able to analyses 1000 sequences at any given time and was used to analyses patient sequences and identify HIV-1 subtypes in the cohort (de Oliveira *et al.,* 2005).

2.2.8.2 Jumping profile Hidden Markov Model Analysis

The jumping profile Hidden Markov Model (jpHMM) (http://jphmm.gobics.de/) method compares query nucleic acid or protein sequences to a family of sequences in a jumping alignment approach as proposed by (Spang et al., 2002). A single query sequence is therefore aligned and compared to different sequences of a single alignment. This approach simplifies the identification of recombinant events as breakpoints of recombinant isolates will be compared to different sequences in the alignment (Zhang et al., 2006; Schultz et al., 2009).

2.2.8.3 Subtype Classification Using Evolutionary Algorithms

Frequent recombination events have lead to the formation of different CRFs and URFs. Although many subtyping software exists to identify new genetic mosaics, few are used to identify inter-subtype recombination. Subtype Classification Using Evolutionary Algorithms (SCUEAL) analysis detects recombination in multiple sequence alignments and identifies breakpoints, recombinant lineages and subtypes http://www.datamonkey.org/dataupload_scueal.php. This computational technique provides a more precise characterization of HIV-1 strains and is particularly useful for identifying novel recombinants. SCUEAL analysis is only available to analyze *pol* sequences for intra and inter clade recombination.

2.2.9 Phylogenetic Analysis using MEGA

2.2.9.1 Multiple Alignments of query and reference sequences

Reference sequences were retrieved from the LANL (http://www.hiv.lanl.gov/) database for the *gag* gene (1237-1721bp, HXB2), the *pol* gene (2136-3335 bp, HXB2), the accessory genes (4954-6228 bp, HXB2) and the *gp41*gene (8321-8709 bp, HXB2) fragments. Reference sequences from the 2010 dataset and query sequences were aligned in Clustal X v2.1 (Thompson et al., 1997). The multiple alignments were converted to Mega format (.meg). All reference sequences were labeled only with the family subtype and GenBank Accession number.

2.2.9.2 Construction of Neighbor-joining trees

Converted MEGA files (.meg) were analyzed in MEGA v5.1 (Tamura et al., 2011) and Neighbor-Joining trees were constructed with Kimura-2 parameter method of distance estimation (Kimura, 1980). Kimura 2 parameter is a method for inferring evolutionary distance in which transitions and transversions are treated separately. The accuracy of the generated tree was validated by bootstrap analysis with total of 500 bootstrap replicates for each dataset, in other words the dataset was resampled and trees redrawn to achieve confidence levels indicated on neighbor-joining trees.

Confidence levels of 70% and greater were considered as significant (Felsenstein, 1985).

2.2.10 Analysis of resistance mutations within patient cohort

The HIV-*pol* gene produces proteins which are pivotal for the maturation of viral progeny, an important step of HIV-1 replication and is therefore an attractive target for ARV therapies. It is also closely associated with drug resistance mutations (DRM) and was therefore cohort sequences were submitted to the HIVdb Program: Sequence Analysis program and analyzed for resistance mutations (http://hivdb.stanford.edu/). The HIVdb program receives query protease and RT sequences, and identifies levels of resistance to 20 commonly used protease and RT inhibitors.

In addition to this *pol* sequences were also analyzed with the Calibrated Population Resistance (CPR) Tool version 6. (http://cpr.stanford.edu/cpr.cgi). CPR uses a list of surveillance drug resistant mutations (SDRM) endorsed by the World Health Organization (WHO) to identify transmitted resistance. The CPR program generates a profile alignment using FASTA-formatted sequence data and compares them to a consensus subtype B sequence. SDRM's are used as markers to identify transmitted resistance in query sequences in the three classes of antiretroviral drugs: PI's, NRTI's and NNRTI's (Gifford et al., 2009).

2.3. Results

2.3.1 Patient demographics

The cohort of 64 patients consisted of both males and females investigated for HAND. A summary of the clinical information and patient demographics are provided in **Table 2.14.** The complete patient information and demographics are listed in **Appendix A.**

Table 2.14: Patient information and demographics

	Variable	Total study population (n=64)
Age	Range	21-36years
	Mean	30years
Gender	Male	17 (27%)
	Female	45 (70%)
	Unknown gender	2 (3%)
HAND status	No HAND	18 (28%)
	ANI	4 (6%)
	MND	26 (41%)
	HAD	11 (17%)
	Unknown	5 (8%)

2.3.2 Nucleic acid extractions

DNA and RNA were isolated from patient PBMCs and plasma respectively. A good DNA concentration level was obtained for all patient samples and ranged between 45.06ng/µl and 192.89ng/µl. The purity levels of all samples were good with most being above 1.80 (260/280).

2.3.3 Polymerase chain reactions and gel electrophoresis

Fifty seven (89%) of the cohort samples were amplifiable for the 484bp fragment of the *gag* gene. However, for the *pol* gene only 47 (73%) of the cohort samples were amplifiable and generated DNA fragments of approximately 1.2 kb in length. Although dithiothreitol (DTT) was added to the PCR reaction to increase specificity, the amplifiability of the *pol* PCR was significantly lower than that of the *gag* PCR and is possibly due to the genetic variability of the cohort and length of the fragments.

For the 1.3kb *accessory* gene fragment, 51 (80%) samples could be amplified while for the 388bp *gp41* gene fragment, 48 (75%) samples could be amplified. A summary of the PCR results are listed in **Table 2.15**.

Table 2.15 Summary of the PCR results for the different fragments

	RNA		**DNA**	
	gag	*pol*	*accessory*	*gp41*
N=sample	64	64	64	64
% Positive	57 (89%)	47 (73%)	51 (80%)	48 (75%)
% Negative	7 (11%)	17 (27%)	13 (20%)	16 (25%)

All the amplified fragments were visualized by gel electrophoresis. A representative gel is presented in **Figure 2.3**., in which 12 partial *gag* fragments were visualized along with a positive control and NTC (non-template control). Samples JO161, JO254 and JO258 could not be amplified, however bright single DNA bands were obtained for the remaining 9 samples.

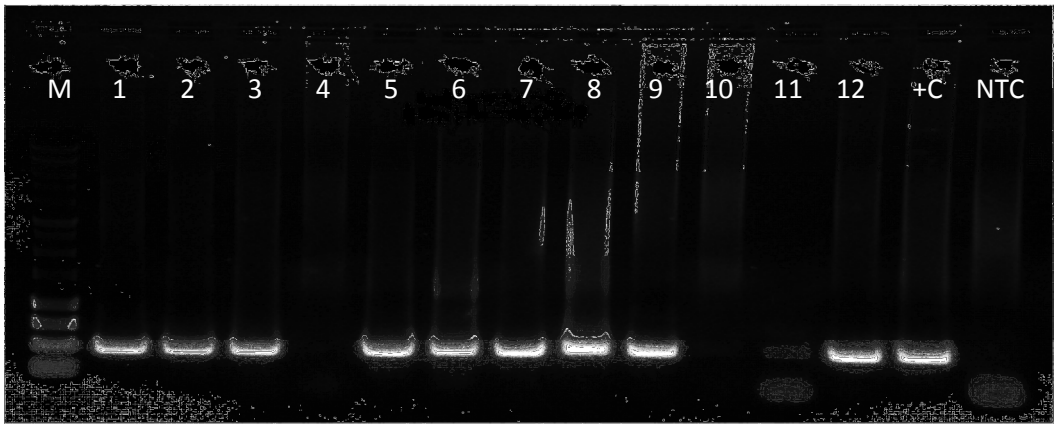

Figure 2.3: Partial *gag* gene fragments amplified in nested PCR assay**.** After two rounds of amplification the 484bp is visualized on a 1% agarose gel. Twelve samples were visualized on the agarose gel along with a 1kb molecular marker (M), lane 1 (JO229), lane 2 (JO230), lane 3 (JO161), lane 4 (JO129C), lane 5 (JO240), lane 6 (JO238), lane 7 (JO248), lane 8 (JO247), lane 9 (JO252), lane 10 (JO254), lane 11 (JO258), lane 12 (JO259), Positive control (+C), and Non-template control (NTC).

2.3.4 Sequence quality control

Sequences quality control identified all *gag* genes as subtype C, except JO162, JO180, JO201 and JO209. A subtype could not be assigned to JO239. Sample JO162 was identified as an A1C recombinant however HIV-1 blast results identified the sample as a subtype C. The samples JO201 and JO209 were both identified as A1 subtypes. HIV-1 blast confirmed JO201 as an A1 subtype with from Sweden however JO209 was identified as a CRF06_cpx from the Niger.

Analysis of *pol* sequences identified JO201 as an A1 subtype, JO209 as a G subtype and JO224 as a C/F1/G recombinant. HIV-1 blast identified AM041053 (subtype A) from DRC, FN599730 (CRF06_cpx) from Senegal and FJ199751 (subtype C) from South Africa as the most closely related sequences for JO201, JO209 and JO224 respectively.

Accessory gene analysis identified 2 samples as possible recombinants. Sample JO207 was identified as an A1/G recombinant with JO403028 (subtype B) from Switzerland as its closest relative. Sample JO286 was identified as a C/H recombinant but HIV-1 blast identified AY13416 from Senegal as its closest relative.

Env gp41 analysis identified JO207 and JO218 as A1 and B subtypes respectively. HIV-1 blast identified HO540689 an A subtype from Kenya and DD153311 a subtype B from France as the closest associated relatives of JO207 and JO218 respectively. JO224 was the only sequence which could not be assigned a subtype. Hyper-mutation was not detected in any of the sequences; sequences identified with frameshifts and stop codons were re-analyzed and verified (Data not shown).

2.3.5 HIV-1 subtyping using online subtyping tools

2.3.5.1 Recombinant Identification Program (RIP)

Sequence analysis of *gag*, *pol*, *accessory* gene and *env* gene fragments identified 7 sequences as non-subtype C or possible recombinants. A summary of RIP results can be found in **Table 2.16**.

Table 2.16 RIP analysis of non-subtype C sequences

	gag	***pol***	***accessory genes***	***env gp41***
JO162	A1/C	-	-	C
JO189	A1/A2/C/D/F1/F2/J	C	C	-
JO201	A1	A1	-	-
JO207	C	-	A1/G	A1
JO209	A1	G	-	-
JO218	C	-	C	B
JO224	C	C/F1/G	C	Not determined
JO286	-	-	C/H	C

2.3.5.2 Rega 3.0 subtyping analysis

The Rega 3.0 tool was used to analyses *gag* p24, *pol*, *accessory* and *env gp41* gene fragments and designated sequences according to HIV-1 subtypes. A total of 52/54 (96%) *gag* sequences were characterized as subtype C. However 2 sequences JO202 and JO247 only showed 81% and 86% support for subtype C characterization respectively. Two (4%) sequences were also characterized as A1 subtypes. *Pol* gene analysis identified 44/47 (94%) sequences as subtype C, 1 sequence as a HIV-1 CRF 06_CPX and 1 sequence as a K subtype, however the support for this subtype designation was only 38%. Accessory gene analysis JO207 was identified as an A1 subtype with 100% support, with the remaining 45/46 (98%) samples subtyped as C's. Env gp41 analysis identified 3 possible recombinants: JO207 (subtype J), JO218 (subtype B), and JO224 (subtype A1). The subtype support was however low for JO207 (33%) and JO224 (45%) while JO218 had a subtype support of 99%. Results of possible recombinants and non-subtype C's have been summarized in **Table 2.17**.

Table 2.17: Summary of possible recombinants and non-subtype C's identified with REGA

	gag	***pol***	***accessory genes***	***env gp41***
JO162	C	-	-	C
JO189	C	C	C	-
JO201	A1	A2	-	-
JO207	C	-	A1	J
JO209	A1	06_cpx	-	-
JO218	C	-	C	B
JO224	C	K	C	A1
JO286	-	-	C	C

2.3.5.3 jpHMM Analysis

Genetic subtyping with jpHMM identified the majority of the cohort as subtype C; however genetic profiling also identified a number of possible recombinants in the patient cohort. JpHMM analysis also identified two additional sequences JO176 and JO213 as possible BC recombinants. The jpHMM results are summarized in **Table 2.16**.

Table 2.16 jpHMM results of non-subtype C sequences

	gag	*pol*	*accessory genes*	*env gp41*
JO162	A2	-	-	C
JO176	C	BC	-	-
JO189	C	C	C	-
JO201	A1	A1	-	-
JO207	C	-	A1/G	A1
JO209	A1	G	-	-
JO213	C	C	BC	C
JO218	C	-	C	B
JO224	C	-	C	C
JO286	-	-	C	C

2.3.5.4 SCUEAL Analysis of the *pol* gene

SCUEAL analysis identified 38/47 (81%) sequences as pure subtype C's, 2/47 (4%) sequences as inter-subtypes (C/G, G/U) and 6/47 (13%) sequences as intra-subtype C's. The JO201 was identified as an A1 ancestral subtype. SCUEAL results are summarized in **Table 2.17**.

Inter-subtype recombinants

SCUEAL identified JO209 as a G/U inter-subtype recombinant with a confidence assignment of 0.81, and support for recombination of 1. In addition JO224 could be identified as a C/G recombinant with confidence assignment of 0.81 and recombination support of 0.96.

Intra-subtype recombination

Sequences JO183; JO186; JO193; JO210; JO232 and JO240 were identified with intra-subtype recombination with either 1 or 2 recombinant breakpoints. Confidence assignments for intra-subtype recombination ranged between 0.53 and 0.90 while support for recombination ranged between 0.74 and 0.99.

Table 2.17 SCUEAL results of recombinants

Name	Subtype	Expanded subtype	Confidence in Assignment	Support for recombination	Support for intra-subtype recombination
JO183	C	C intra-subtype recombinant (2 breakpoints)	0.903324	0.90339	0.90339
JO186	C	C intra-subtype recombinant (2 breakpoints)	0.776941	0.777905	0.777905
JO193	C	C intra-subtype recombinant (1 breakpoints)	0.530475	0.744084	0.744084
JO201	A-ancestral	A-ancestral	0.988561	0.000586802	0
JO209	G,U recombinant	G,U inter-subtype recombinant.	0.813492	1	0.000175057
JO210	C	C intra-subtype recombinant (1 breakpoints)	0.656377	0.948072	0.948069
JO224	C,G recombinant	G,C inter-subtype recombinant.	0.815037	0.964625	0.00205636
JO232	C	C intra-subtype recombinant (1 breakpoints)	0.823584	0.931891	0.93189
JO240	C	C intra-subtype recombinant (2 breakpoints)	0.715358	0.991571	0.991571

2.3.6 Phylogenetic Analysis using MEGA version 5.1

Phylogenetic analysis was used to investigate the relationships between cohort sequences and subtype reference sequences and also to confirm the results obtained from online subtyping tools.

Construction of partial *gag* Neighbor-joining tree

A neighbor joining tree (**Figure 2.4**) was constructed using sequences from the LANL database [http://www.hiv.lanl.gov]. The analysis involved 81 nucleotide sequences. The majority of the samples clustered with subtype C sequences: C.AF067155, C.AY772699, C.U46016, CU52953 (indicated in **Figure 2.4B**), however JO201 was identified as an outlier and clustered around A1 sequences: A1.DQ676872, A1.AB253429, A1.AB253421, and corresponded with the Rega 3.0 and jpHMM analysis. In addition JO209 clustered with CRF06_cpx.AF064699 and CRF06_cpx.AY535659; however jpHMM and Rega 3.0 analysis identified this sequence as an A1 subtype.

Construction of partial *pol* Neighbor-joining tree (Figure 2.5)

The phylogenetic clustering observed in the *gag* Neighbor Joining tree was similar to the clustering of the *pol* tree. The same clustering patterns were observed with subtype C and the reference sequences C.AF067155, C.AY772699, C.U46016, and CU52953 (indicated in **Figure 2.5B**), as well as with JO201 and the A1 sequences: A1.DQ676872, A1.AB253429, A1.AB253421, and JO209 with CRF06_cpx.AF064699 and CRF06_cpx.AY535659 (**Figure 2.5A**). The similar clustering patterns observed in the *gag* and *pol* trees therefore verify the results obtained.

Construction of Accessory gene Neighbor-joining trees (Figure 2.6)

The *accessory* phylogenetic tree identified four cluster groups with the largest consisting of 21 query sequences and C.AY772699. Smaller clustering was observed between JO183, JO195, JO286 and C.U52953. JO213 was identified as an outlier while JO207 clustered near A1.DQ676872 and A2.AF286238. Due to the short DNA sequence generated for gp41 a reliable tree could not be generated.

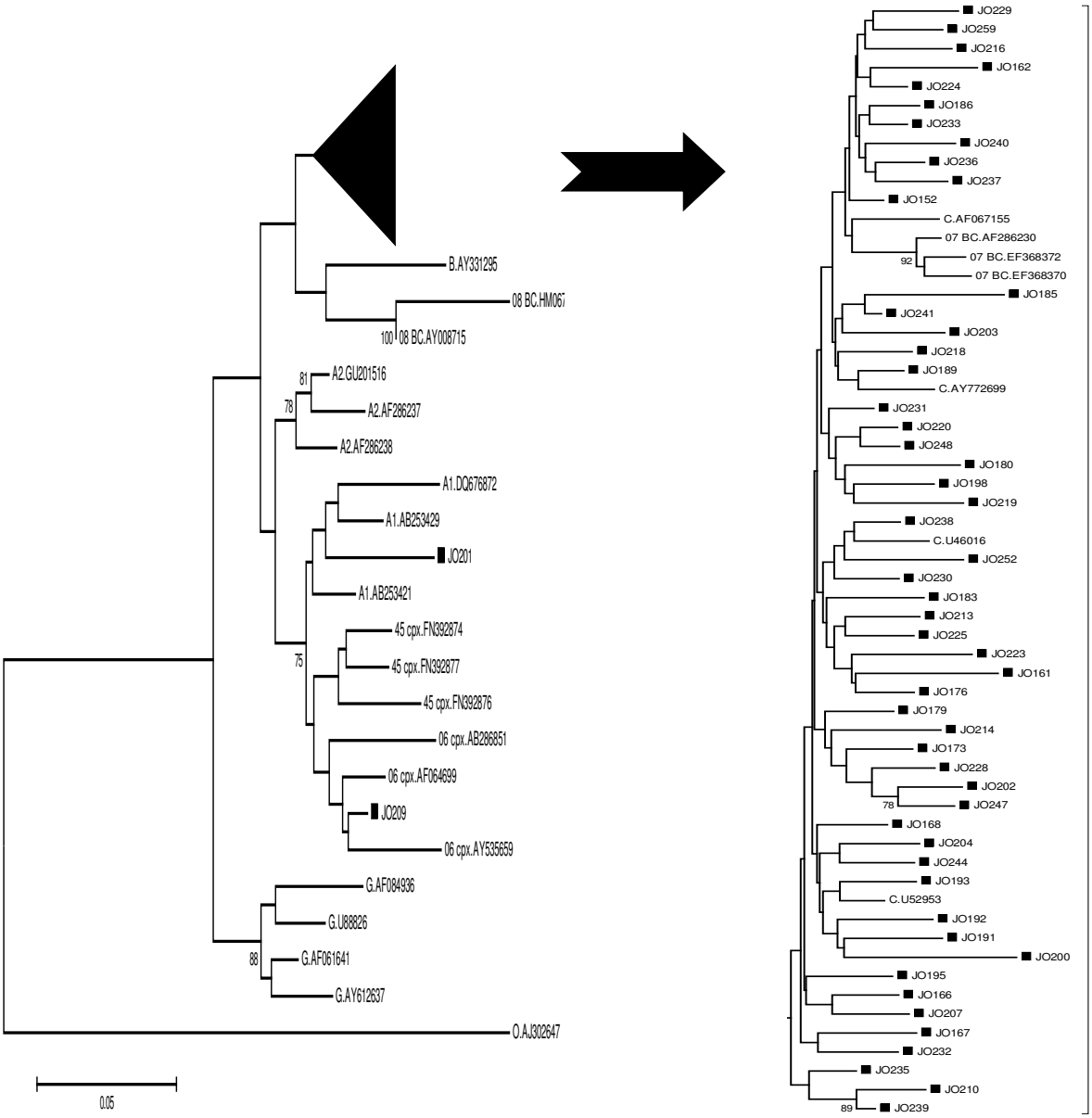

Figure 2.4 Neighbor joining tree of *gag* gene. A Neighbor joining tree of *gag* sequences (~ 484bp) and reference sequences was constructed. Each query sequence is indicated with black squares. Bootstrap values of greater than 70% are indicated on the condensed tree (A) and uncondensed (B) portion of *gag* tree.

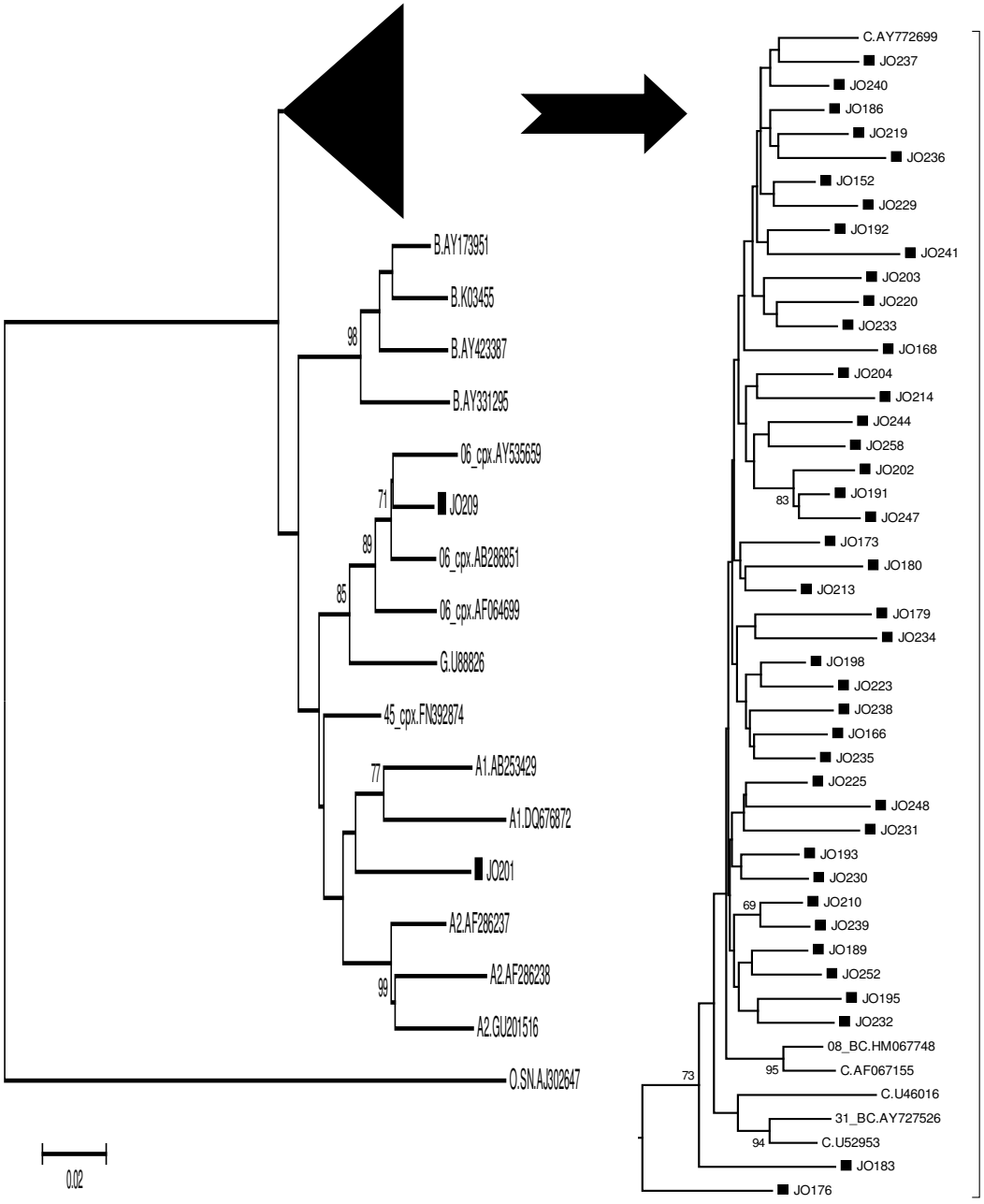

Figure 2.5 Neighbor joining tree of *pol* gene. A Neighbor joining tree of *pol* sequences (~1.2kb) and reference sequences was constructed. Each query sequence is indicated with black squares. Bootstrap values of greater than 70% are indicated on the condensed tree (A) and uncondensed (B) portion of *pol* tree

Figure 2.6 Neighbor Joining tree of *accessory* reference and query sequences. The phylogenetic tree was drawn in Mega 5v.1 with Kimura 2 parameter and consisted of cohort and query sequences approximately 1.3kb in length. Bootstrapping was performed with 500 replicates with significant values indicated on tree. Query sequences are indicated with colored squares.

2.3.7 Analysis of resistant mutations within the patient cohort

Analyses of antiretroviral resistance were primarily based on HIV RNA extracted from patient plasma. The *pol* sequences were used for the identification of possible major and minor mutations in the three antiretroviral categories namely: PIs, NRTIs and NNRTIs. Sequences were screened for mutations using the Stanford HIV-1 drug resistance database as well as CPR. The CPR tool could not identify any transmitted resistance in any of the sequences. The Stanford HIV-1 database detected no viral susceptibility to antiretroviral compounds in most *pol* sequences. However eight samples: JO189, JO209, JO213, JO214, JO220, JO225, JO232 and JO233 all displayed low level resistance to Nelfinavir (NFV), a PI. However, most of these sequences contained no minor or major PI mutations or NRTI and NNRTI mutations with the exception of JO209.

In addition JO209 and five other sequences: JO186, JO201, JO236, JO247, JO248 could all be identified with either a minor PI mutation or a NNRTI mutation. JO186 and JO201 were both identified with L10I, a minor PI mutation while JO248 was identified with another minor PI mutation A71T. JO209, JO236, JO247 were identified with the NNRTI mutations: V179E, K103R, E138A respectively.

2.4. Discussion

Genetic sequencing investigates the nucleotide base sequence of an organism and utilizes subtyping programs to characterize viral strains. However genetic characterization can prove difficult as a result of high genetic variability (Spira et al., 2003). It is therefore important to utilize different programs to characterize the virus. In this study g*ag*, *pol*, *accessory* gene and *gp41* gene fragments were analyzed with subtyping programs (RIP, Rega 3.0, jpHMM and SCUEAL) to characterize the viral subtypes in the HAND cohort and confirmed with phylogenetic trees. In addition to vial subtyping *pol* sequences were also analyzed for possible drug resistance.

2.4.1 Sequence Analysis of HAND cohort

Using Rega 3.0 and jpHMM subtyping tools, we collectively identified two possible B/C recombinants (JO176, JO213), one A1 subtype (JO201), two A1/G recombinants (JO207 and JO209) and one A1/C recombinant (JO224).

Subtype B has been reported in parts of Asia, Brazil, India and America. However, the co-circulation of subtype B and C in IDU s in Brazil and Asia have resulted in two recombinant forms CRF07_BC and CRF08_BC. Similarly, subtype B is the dominant strain in Western Europe, Australia and North America; however the diversity of subtype B is greatest in South American countries such as Brazil and Argentina with the circulation of B, C, F, BC and BF recombinants (Tebit & Arts, 2011). Likewise the co-circulation of subtype B and subtype C in heterosexual and homosexual populations in South Africa could explain the BC recombinants identified in the HAND cohort. Conversely ambiguous results were obtained for JO176 and JO213.

Cohort sequences were further analyzed with SCUEAL, which allowed for more in depth analysis of subtype classification, especially for sequences which were previously thought to be pure single subtypes. Apart from subtype C, subtype A1 was the most common amongst the recombinants and was correctly detected by all three online sequencing tools. The number of recombinants identified initially expanded from the six inter-subtype recombinants to include an additional nine cases of intra-subtype C recombinants, drastically increasing the level of recombination from 9.3% to 23.43%. Interestingly, recombination occurred in 20.31% of HAND patients with recombinants being identified in 3 cases of ANI (4.68%), 8 cases of MND (12.5%) and 2 cases of HAD (3.12%). It is clear that a trend towards more complex recombinants is emerging, but it is still unclear whether this level of recombination is significant to HAND patients.

Similarly, HIV-1 subtype C was identified as the dominant strain in the HAND cohort, with a number of subtype C clusters observed in both *gag* and *pol* phylogenetic trees. This finding indicates that this subtype population within the HAND cohort did not originate from a single common ancestor but rather is due to the introduction of multiple viruses from various geographical sources. The distribution of the subtype C strain is not confined to sub-Saharan Africa and has established itself in a number of other countries. Subtype C circulates at low levels in Kenya and Uganda as either a pure subtype or as a recombinant strain, but with decreasing incidence. Previous reports state that the circulation of subtype C in China is a by-product of the introduction of subtype C from South Africa to India. Other studies have also

identified an association between the subtype C epidemic in East Africa, South America and the United Kingdom (UK) (Tebit & Arts, 2011; de Oliveira et al., 2010).

2.4.2 Phylogenetic analysis of HAND cohort

The clustering pattern observed for JO176 and JO213 was identified them as subtype C in both the gag and accessory phylogenetic trees. For JO213 the *pol* phylogenetic tree clustered with subtype C strains while the clustering pattern for JO176 suggested that it was indeed a BC recombinant. Though, due to the ambiguity of these samples, full genome sequencing would be necessary to fully elucidate the subtype of these samples.

The gp41 and accessory genes of JO207 was characterized as an A1 recombinant and clustered with sequences originating from Rwanda and Australia. This clustering pattern is unusual and could have been introduced with immigration. The dataset used to characterize JO207 also included another African A1 sequence, which originated from Uganda. Subtype A has been a major circulating strain within Uganda and Kenya since the mid-1980's, (Hu et al., 2000), with the Kenyan subtype A strain sharing homology with other African countries such as Uganda, South Africa, Ethiopia, Rwanda, Botswana and Democratic Republic of Congo (DRC). This may indicate direct transmission between migrants. Subtype A has also been slowly evolving and diverging into new strains which share homology with strains in India, Australia and Sweden, possibly indicating the spread of these strains from or to Kenya (Khoja et al., 2008).

The sequence for patient JO209 was screened for recombination and could not be easily subtyped. Phylogenetic analysis identified the sample as a CRF06 with a mosaic genome structure consisting of subtypes A, G, J and K. This is the first identification of the virus in South Africa, which originates from the Congo, and indicates the introduction of new CRF's into the South African population (Zhang et al., 2010).

The data of this cohort was correlated with previous studies, which found subtype C as the dominant strain in Cape Town. However, frequent detection of recombinants and URF's has made the HIV-1 epidemiology in Cape Town progressively more heterogeneous. According to Neogi et al, pure subtypes are steadily being replaced

by more complex mosaic genome structures (Neogi et al., 2011). This is evident in the increasing number of recombinants detected in recent years. In addition, a study conducted by Jacobs et al., (2008) found that a small percentage of their cohort were resistant to the current ART regime. Other studies have observed a similar trend in treatment naïve patients with 10% identified as being resistant at baseline.

2.4.3 Analysis of Drug Resistance Mutations in Cohort

In 2001 an ART roll out program was initiated in areas of Cape Town. By 2008 approximately 213 000 people had been enrolled and were receiving ART on a daily basis. ART therapy works by suppressing HIV-1 replication through intercepting viral RT and PR activities (Clavel and Hance 2004). However, drug therapy and immunological pressures also actively selects for drug resistant viral strains. These strains are then archived in resting white blood cells and re-emerge once ART fails and can be spread within the population (Shafer et al., 2007).

During this study HIV-1 sequences from treatment-naive patients were analyzed for possible resistance mutations. Low level resistance to the PI NFV was observed in 8 cohort samples. According to the Stanford database (http://hivdb.stanford.edu), drug resistance to this inhibitor is caused by the presence of mutations such as D30N working in conjunction with other accessory mutations: L63P, A71V and L90M. NFV resistance can also occur in the presence of T74S, a mutation which improves viral fitness in subtype C strains. The T74S mutations is reported to have a high natural mutation frequency in treatment-naïve subtype C patients of 18% (Soares et al., 2009). Our cohort had a T74S mutation frequency of 14.06% with all 8 samples being identified with this mutation and had a mutation score of 15 for the NFV inhibitor, which according to the Stanford resistance database, is indicative of low-level resistance. In addition the observed NFV resistance occurred in 4 patients (JO209, JO189, JO214 and JO220) diagnosed with MND. According to Tozzi et al., (2007), patients with advanced cognitive impairment have a higher risk of developing ART drug resistance. Even more concerning is the fact that these patients are treatment naïve which may pose a problem for upcoming treatment.

2.5. Conclusion

This study has shown that sequences of the HAND patients are predominantly subtype C with at least 15 cases of possible inter-subtype and intra-subtype C recombinants being identified. In addition, epidemiological trends are difficult to predict, especially since the patterns of internal and cross-border mobility have changed considerably over the last three decades and may therefore contribute to the current epidemiological shift of HIV-1 being observed in South Africa. Continued evolution of subtypes and emergence of new URFs and CRFs may further complicate the development of effective strategies against HIV-1and therefore emphasizes the need for continued HIV-1 surveillance. Furthermore, the presence of resistance in our cohort at baseline highlights the necessity of drug resistance testing surveillance especially for treatment naïve patients.

CHAPTER THREE

Molecular characterization of the *tat* gene and protein

3.1. Introduction

In a recent article, Cowley et al (2010) identified a number of point mutations (Chapter 1, **Figure 1.8**) within HIV-1 Tat subtype B isolates. These mutations were identified within CNS and lymphoid tissue and were localized within the first exon of Tat. The majority of the mutations were associated with the transactivation ability of the *tat* gene. According to Cowley et al. (2010) these mutations disrupts the RNA stem loop structure (TAR) and reduces *tat* transactivation.

Primarily, this chapter will focus on HIV Tat exon 1 as limited amino acid composition and point mutational information is available for Tat exon 2. The chapter will also investigate genetic conservation of full-length, cohort-derived Tat sequences and characterize the amino acid composition of the Tat exon 1 presented in the cohort. This characterization allows us to determine the homology between the cohort-derived tat exon 1 sequences and those described by Cowley et al. (2010) found in the CNS; it also affords us the opportunity to determine the frequency of these amino acids in subtype B and subtype C sequences (obtained from the LANL database); and, finally, provides some insights into the amino acid signature patterns of the chemotactic and neurotoxic domains of Tat exon 1. A schematic overview of the sequence databases and software packages used to characterize the *tat* gene and protein is shown in **Figure 3.1.**

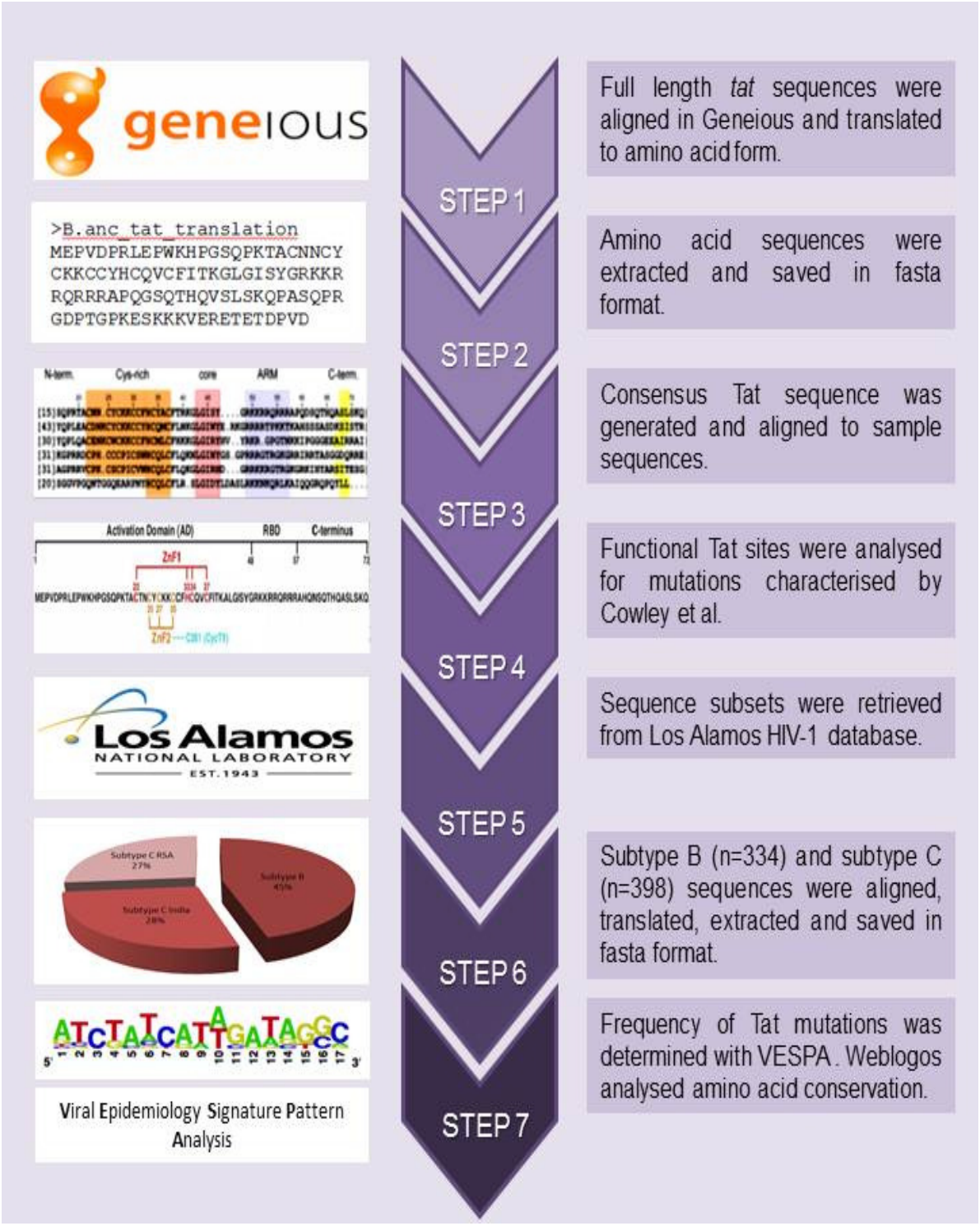

Figure 3.1 Overview of steps used to analyze HIV-1 *tat* genes and Tat proteins.

3.2. Materials and methods

3.2.1 Software and databases used for signature pattern analysis

All the reagents, equipment, and software applications that were used during the course of this study are listed in **Tables 2.1 - 2.3**. Registered trade mark items and trade mark products are indicated by the symbols ® and TM respectively.

Table 3.1. List of chemicals and kits used during the study.

Software packages	References and/or licensed companies
Geneious 5.5.7	Biomatters Ltd, Aucklandland, NZ
SeqPublish	http://www.hiv.lanl.gov/content/sequence/SeqPublish
Los Alamos	http://www.hiv.lanl.gov/components/sequence/HIV
VESPA	http://www.hiv.lanl.gov/content/sequence/VESPA
WEBLOGOS	(Computational Genomics Research Group, University of California, Berkley)

3.2.2 HIV-1 tat sequence datasets

Three sequences datasets were retrieved from the LANL database (http://www.hiv.lanl.gov/components/sequence/HIV/search/search.html) and used to compare the various mutations present in the first exon of the *tat* gene. These sequence datasets included: Indian subtype C (n=203), South African subtype C (n=195) and subtype B sequences (n=334).

3.2.3 Nucleotide and amino acid sequence alignments

The study cohort sequences were imported into Geneious 5.5.7 (Biomatters Ltd, New Zealand). Subtype C cohort sequences and the HIV-1 *tat* exon 1 and HIV-1 *tat* exon 2 gene fragments were extracted. Complete *tat* gene sequences were then assembled, realigned and translated into amino acid format.

Subtype B and Subtype C (Indian and South African) sequences datasets obtained from the Los Alamos Database, were aligned and the alignments were used to generate a consensus sequence for each dataset. The sequence alignment publisher tool, SeqPublish was used for this. (http://www.hiv.lanl.gov/content/sequence/SeqPublish/seqpublish.html). The interface uses a sequence alignment (nucleotides or amino acids) and identifies similarity between sequences with dashes. The reference sequence is either the first sequence in the alignment, or the software can create a consensus sequence to use in subsequent analyses.

3.2.4 Signature Pattern Discovery and Recognition

Signature pattern analysis is rapidly becoming an essential tool for genetic sequence analysis. The analysis software utilizes specific clusters of residues or “patterns” to identify proteins which are functionally similar; a task which conventional sequence alignment and mapping software is unable to perform (Hulo et al., 2006). Recent studies, focusing on the analysis of host proteins incorporated by lentiviruses, has highlighted the importance of amino acid characterization, emphasizing its potential role in better understanding HIV-1 pathogenesis and the prospects for the identification of new ARV therapy targets (Li et al 2012). Signature patterns between background sequences and query sequences were analyzed using two bioinformatics software packages: (1) viral epidemiology signature pattern analysis (VESPA) and (2) WebLogos (Computational Genomics Research Group, University of California, Berkley). Each tool was used to analyses specific aspects of the tat sequences resulting in a comprehensive overview of the observed conservation in cohort-derived tat proteins at a molecular level.

VESPA

VESPA analyzed the *tat* motif and point mutations in *tat* exon one only, as mutations in the second *tat* exon have not yet been functionally characterized. VESPA defines signature patterns, such as atypical amino acid or nucleotide residues, and calculates the frequency of each amino acid or nucleotide at each position in an alignment of two sequence subsets. The VESPA program identifies mutual amino

acids, conferring altered functionality, as well as non-functional differences between the reference and the query sequences.

Three types of patterns can be identified with VESPA: (1) an inflexible signature in which uncommon residues in the background sequence are also present in query sequences, (2) dominant residue signatures in which the more common residues in the query sequences differ from that seen in the reference sequence, (3) direct signatures in which a consensus is generated and used to reflect the predominant amino acid at each position (Korber and Myers, 1992).

WebLogos

WebLogos exclusively interrogated the neurotoxic and chemotactic domains of Tat, spanning amino acids 24–51 and 31–61 respectively. Characteristically, Tat displays a great neurotoxin and chemotactic affinity. Any amino acid substitutions in this domain are likely to affects both mechanisms.

Sequences WebLogos are graphical representations depicting the prevalence of an amino acid or nucleic acid multiple sequence alignment. The logos are comprised of stacks of characters with every stack representing a single nucleic or proteomic loci interrogated. Sequence conservation is indicated by the overall height of the character, while the height of characters within the stack indicates their relative frequency at that position. A sequence logo provides a more descriptive representation of the patterns within a multiple sequence alignments and can rapidly reveal significant features of the alignment, which would normally be difficult to recognize.

3.3. Results

3.3.1 Generation of Datasets

The subtype B sequence dataset was the largest amongst the three datasets and included sequences from a total of 33 European, African and Asian countries. Full length *tat* subtype B (*n*=332) and South African subtype C sequences (*n*=179) could be retrieved for comparison to the study cohort. A limited amount of full length *tat* Indian subtype C sequences (*n*=7) could be retrieved from the Los Alamos database.

As a result additional Indian subtype C sequences were analyzed, which consisted only of Tat exon 1 sequences (*n*=203). Sequence datasets are summarized in **Table 3.1** (the complete list of sequences used for this analysis has been described in the **Appendix**). The 3 different sequence datasets were compared to full length *tat* sequences from the patient cohort (*n*=34).

Table 3.1 Characterization of HIV-1 sequence datasets

Tat subtype B and subtype C sequence datasets					
Code	**Name of country**	**Number of taxa**	**Code**	**Name of country**	**Number of taxa**
AR	Argentina	9	HT	Haiti	4
AU	Australia	29	**IN**	**India – subtype C**	**210**
BO	Bolivia	1	IT	Italy	1
BR	Brazil	18	JM	Jamaica	1
CA	Canada	6	JP	Japan	15
CN	China	4	KR	Korea, Republic of (South)	4
CO	Colombia	4	MM	Myanmar	1
CU	Cuba	2	NL	Netherlands	1
CY	Cyprus	32	RU	Russian Federation	3
DE	Germany	1	TH	Thailand	3
DK	Denmark	13	TT	Trinidad and Tobago	4
DO	Dominican republic	2	TW	Taiwan, Province of China	1
EC	Ecuador	2	UA	Ukraine	3
ES	Spain	7	US	United States of America	147
FR	France	1	UY	Uruguay	2
GA	Gabon	1	YE	Yemen	2
GB	United Kingdom	3	**ZA**	**South Africa – subtype C**	**179**
GE	Georgia	2	**TOTAL**		721
HK	Hong Kong	1			

3.3.2 Sequence alignments

A consensus subtype C sequence was generated from the study cohort and aligned to the cohort sequences in **Figure 3.2** to determine the level of amino acid variability in HIV-1 tat. The alignment focuses on the various domains of tat and indicates the level of variability across these five domains. The observed point mutations showed clustering in the cys-rich domain, basic domain and parts of the C-terminal domains.

Bioinformatics was utilized to investigate the landscape of the *tat* gene and to obtain comparative data regarding the genetic and amino acid variation in the HAND cohort. VESPA and WebLogos were utilized to identify areas of conservation and hotspots of variability in tat. .

A number of sequence specific amino acids could be identified in the HAND cohort with the greatest number of polymorphisms being observed in the cysteine rich, glutamine-rich and C-terminal domains **(Figure 3.2).** Furthermore mutations were also observed in the overlapping areas of the chemotactic and neurotoxic domains of Tat. The sequences JO193 and JO204 were identified with the least amount of variability but harbored mutations at H29K, A58T, P68L, D80N, P81S and K19Q, K24T, H29K, P81Q, E96K respectively. Furthermore, the sequences labeled: JO168, JO218 and JO235 presented with premature stop codons with JO168 and JO218 containing stop codons at F100* and JO235 at position S86*. It has been shown that premature stop codons lead to truncated protein forms which are often inadequate for proper cell functioning (Jeang et al., 1999). Although, observed stop codons occurred in the C-terminal domain and may therefore still possess normal functioning, albeit reduced.

Sequence analysis also identified 16 sequences with mutations at sites which have previously been characterized as functionally important for Tat transactivation and post translation modifications such as phosphorylation (Cowley et al. 2010). Mutations associated with transactivation occurred at E2D, C31S/A while mutations associated with phosphorylation was identified at S46Y, S57R, S62G and D64N/G. An unexpected C31A mutation was identified in the dicysteine motif of Tat which has previously not been identified in subtype B sequences. Furthermore, JO207 was identified with the highest level of variability with a total of 26 mutations, 3 of which were located in sites associated with transactivation and phosphorylation. The identification of amino acid changes in functionally important sites reflected the heterogeneity of Tat sequences. Therefore to assess the relative frequency of these mutations an examination of clade B and clade C sequences were performed with VESPA.

3.3.3. Signature pattern analysis of amino acid residues

Sequence WebLogos and VESPA analysis was performed on subtype B and subtype C sequences to identify point mutations in sixteen functionally important sites of Tat. **Table 3. 2** summarize the frequency of conserved amino acids as well as its amino acid substitutions. In addition, the chemotactic and neurotoxic domains

of Tat were analyzed for signature patterns and presented with sequence WebLogo in **Figure 3.3.**

The conservation of amino acids in 13 of the 16 sites was above 80% while the remaining 3: C31S, 57S, and 64D had frequency levels of 50.65%, 46.21% and 53% respectively. Observed amino acid substitutions could be characterized into 1 of 6 categories: Mutations shared amongst all sequence subsets and the cohort, mutations unique to subtype B, mutations unique to subtype C India, mutations unique to subtype C South Africa, mutations shared by subtype B and subtype C and mutations shared between subtype C sequences. Subtype specific mutations were most prevalent in subtype B sequences (22), which were closely followed by South African subtype C sequences (21) and finally subtype C Indian sequences. The frequencies of these subtype specific mutations were between 0.13% and 0.91%. Mutations shared amongst sequences yielded similar frequencies except in positions: E2D, C31S/A/T/N/G, S57R and D64T with mutations frequencies between 15.27% and 48.56%.

Position 31 was of particular interest as mutations here have been implicated in the apoptotic ability of HIV-1 Tat and the progression of HAND. The patient cohort showed similarity to both the subtype B and subtype C strains as it contained both cysteine and serine residues in position 31. Mutation frequencies were as follows: C31S (50.65%), C31C (48.30), C31A (0.52%), C31T (0.26%), C31N (0.13%) and C31G (0.13%). Interestingly the novel C31A mutation identified in the HAND cohort could only be identified in South African subtype C sequences. Sequence WebLogos identified 3 amino acid sequences which were relatively conserved amongst sequence subsets: $_{25}$CYCK$_{28}$, $_{36}$VCF$_{38}$ and $_{41}$KGLGISYGRKKRRQRR$_{56}$ with variability occurring at positions 24, 26, 29, 31, 32, 36, 40, 41 and 57-61. The amino acids Asn, Thr, and Lys were found at position 24 amongst all sequences subsets as well as Pro in Subtype B sequences. Tyr and Phe were recognized at positions 29 however, the latter displayed a greater level of conservation in all four sequences than Tyr.

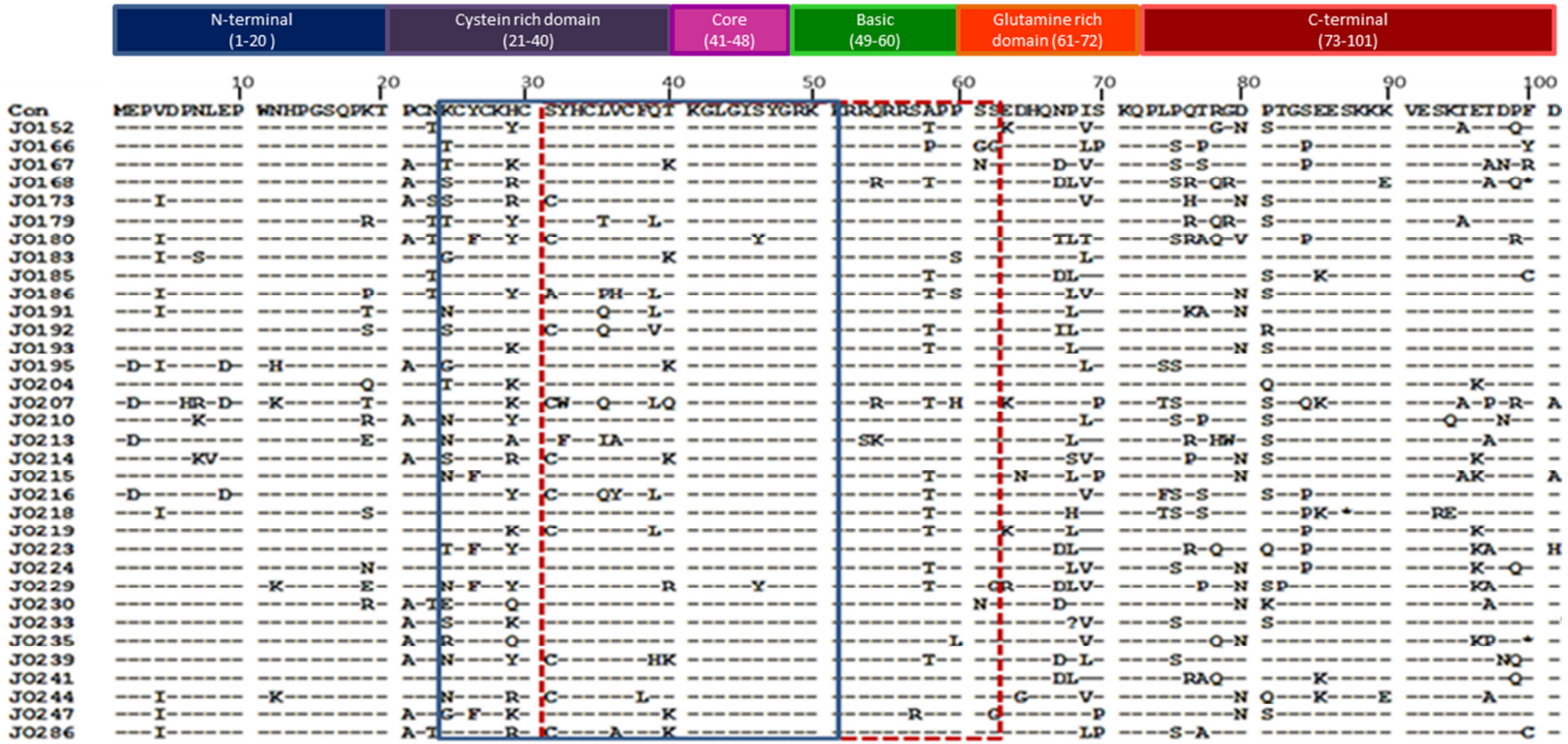

Figure 3.2: Nucleotide sequence alignment of subtype C consensus sequence and 34 full length *tat* sequences. N-terminal, cysteine-rich, core, basic and C-terminal domains are indicated, with blue and red boxed areas representing chemotactic and neurotoxic domains respectively. Dashed areas indicate areas of similarity.

Table 3.2: Frequency of mutations in functionally important Tat protein sites used for transactivation and phosphorylation. **Purple** = Indian subtype C specific mutations, **Green** = South African subtype C specific mutations, **Pink** = subtype B specific mutations, **Red** = mutations identified in subtype C, subtype B and cohort sequences, **Orange** = mutations identified in subtype B and subtype C sequences (India and South Africa), **Turquoise** = subtype C specific mutations, **Yellow** = subtype B and South Africa subtype B specific mutations.

Amino Acid position		Consensus	Polymorphism and relevant frequency											
2	a.a	E	D	G	K	P	R							
	n	695	117	7	2	1	1							
	%	90.73	15.27	0.91	0.26	0.13	0.13							
11	a.a	W	*	S	G									
	n	762	2	1	1									
	%	99.47	0.26	0.13	0.13									
16	a.a	S	N	C										
	n	764	1	1										
	%	99.73	0.13	0.13										
22	a.a	C	R	S	W	A	V							
	n	758	4	1	1	1	1							
	%	98.95	0.52	0.13	0.13	0.13	0.13							
28	a.a	K	N	R	Q									
	n	761	3	1	1									
	%	99.34	0.39	0.13	0.13									
31	a.a	S	C	A	T	N	G							
	n	388	370	4	2	1	1							
	%	50.65	48.3	0.52	0.26	0.13	0.13							
34	a.a	C	G	W	V	R	A							
	n	757	3	2	2	1	1							
	%	98.82	0.39	0.26	0.26	0.13	0.13							
41	a.a	K	Q	N	S									
	n	762	1	2	1									
	%	99.47	0.13	0.26	0.13									
46	a.a	S	Y	F	P	C	K	I	V					
	n	707	34	11	5	4	1	1	1					
	%	92.29	4.43	1.43	0.65	0.52	0.13	0.13	0.13					
47	a.a	Y	H	N	M	G	F	*						
	n	738	15	7	3	1	1	1						
	%	96.34	9.79	0.91	0.39	0.13	0.13	0.13						
50	a.a	K	R	I										
	n	762	3	1										
	%	99.47	0.39	0.13										
51	a.a	K	S	R	N	E	F							
	n	758	2	2	1	1	1							
	%	98.95	0.26	0.26	0.13	0.13	0.13							
56	a.a	R	S	Q	E	A	P	K						
	n	747	35	11	3	1	1	1						
	%	97.51	4.56	1.43	0.39	0.13	0.13	0.13						
57	a.a	S	R	G	T	N	A	K	*	P				
	n	354	372	13	9	8	5	2	1	1				
	%	46.21	48.56	1.69	1.17	1.04	0.65	0.26	0.13	0.13				
62	a.a	S	N	G	R	C	H	D	T	V	Y	E	Q	
	n	658	39	28	15	7	6	4	3	3	1	1	1	
	%	85.9	5.09	3.65	9.79	0.91	0.78	0.52	0.39	0.39	0.13	0.13	0.13	
64	a.a	D	T	N	A	G	I	K	P	Q	S	H	V	E
	n	406	232	45	21	18	17	5	4	2	2	2	2	1
	%	53	30.28	5.87	2.74	2.34	2.21	0.65	0.52	0.26	0.26	0.26	0.26	0.13

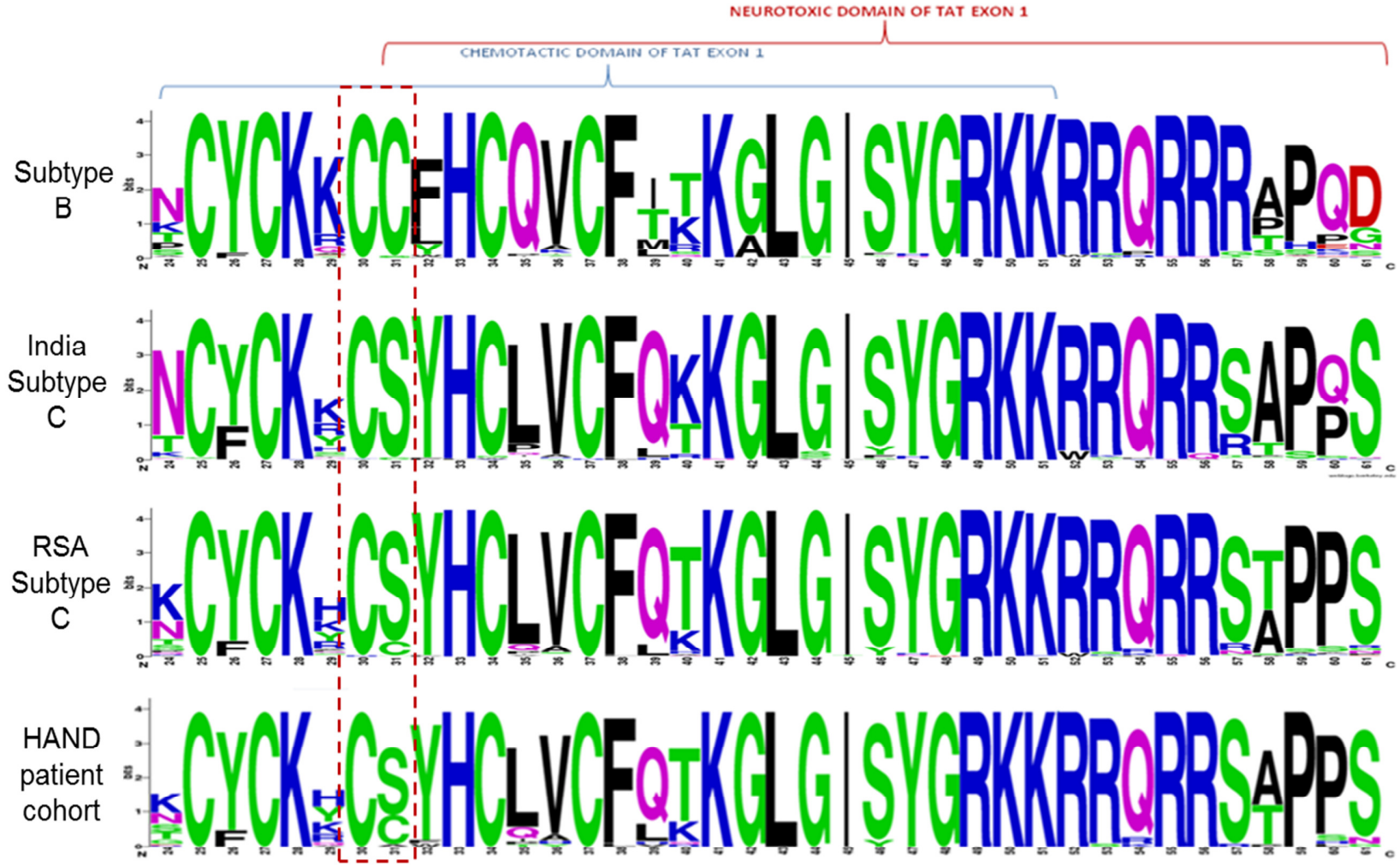

Figure 3.3: Chemotactic and neurotoxic regions of HIV-1 Tat. (A) Subtype B, (B) India Subtype C, (C) South African Subtype C, and (D) cohort sequences. The x-axis indicates the position of amino acid, y-axis indicates the frequency of amino acid. The sequence logo was developed at http://weblogo.berkeley.edu/logo.c

3.4. Discussion

The HIV-1 genome is a hive of constant change and is incessantly subjected to evolution in the form of mutations which are possibly damaging to viral fitness and replication. However the error prone nature of HIV has resulted in a multitude of diverse strains which are not only functional but capable of successfully infecting and replicating in various host cells (Evans et al., 2009). Subtype C viruses predominates in countries such as India and South Africa and are genetically distinct from other subtypes such as HIV-1 B (Ranga et al., 2004). Though, little is known about the genetic differences between HIV *tat* subtype B and subtype C. This chapter sought to identify amino acid differences in Tat, between the viral subtypes B and C. A total of 732 subtype B and subtype C sequences were retrieved from the Los Alamos sequence database. Sequences were analyzed using bioinformatics to identify amino acid which were subtype specific or commonly shared amongst subtypes.

3.4.1 Variability in functionally important amino acid positions

Analysis of Tat amino acids revealed a high level of heterogeneity in both subtype B and subtype C sequences. The significant variation in Tat amino acids was associated with transactivation and phosphorylation. A number of the mutations observed were in agreement with those described (Cowley et al., 2010) with frequency levels of between 80% and 99%. However the sequences did vary greatly in positions, 31, 57 and 64. Cowley identified the mutations as C31R, R57G/K, T64N/D, while in our cohort and the majority of sequences analyzed, the mutations were identified as C31S/A/T/N/G, S57G/T/N/A/K/*/P, D64T/N/A/G/I/K/P/Q/S/H/V/E. Although the level of variability observed in mutations were much greater than those described by Cowley the mutation frequency was often below 1%. The observed variability could exist as a result of HIV-1's erroneous nature, HAART selective pressure or natural selection of subtype C for a more advantageous life cycle.

3.4.2 The Dicysteine Motif

While 99% of non-subtype C viruses can be identified as having a cysteine at position 31, about 4.5% of subtype C sequences in the database encoded for it. The cysteine rich domain of Tat is said to govern the transactivation domain of Tat and contains 7 conserved cysteine, mutations in this area would therefore be detrimental

for HIV-1 transactivation and replication. The study cohort sequences contained no mutations in the 7 conserved cysteine positions, except for position 31. According to (Ranga et al., 2004), the cysteine in position 31 is irrelevant for transactivation in subtype C, but is conserved in non-subtype C's for monocyte chemotactic function, an activity absent in subtype C.

Studies conducted by Mishra et al., 2004 provided evidence that the C30S31 mutation in subtype C decreased chemotactic function without affecting transactivation. Furthermore, the study also showed that the C30S31 mutation was conserved among subtype C sequences regardless of geographic distribution. Conservation of the C30S31 mutation was observed in the analysis of the study cohort as well as subtype C Indian and South African sequences retrieved from the Los Alamos database. Ranga et al (2004) has suggested that the conservation of serine in position 31 amongst subtype C sequences could possibly be functionally advantageous to the subtype, but is yet to be described. Furthermore a study showed that serine residues in subtype C were phosphorylated by a cyclin dependent kinase-2 mechanism (Tiwari et al., 2012) which is important for HIV-1 transcription and the activation of integrated HIV-1 provirus which further supports Ranga's claim that the C30S31 mutation could be beneficial to subtype C and is a possible phosphorylation site which supplements the loss of a cysteine and its trans-activation activity.

In addition, protein studies conducted by Mishra et al., 2008 identified subtype B proteins with C30C31 mutations with a higher order of flexibility than the isogenic form containing the C30S31 mutation. It was suggested that the higher order of flexibility in subtype B could result in the higher level of activity reported for subtype B. This could possibly explain the high conservation of cysteine residues in position 31, observed in the subtype B sequences analyzed in this study. Despite the error prone nature of HIV-1 replication, and the variability observed, Tat still retains its complex functions of gene expression and replication. It has been reported that short conserved amino acid sequences associated with Tat functioning are utilized for this function (Dey et al., 2012).

3.4.3 The chemotactic and neurotoxic domains of Tat

Findings from the sequence alignments prompted further investigations of the patient cohort. WebLogo software provided a detailed and more precise description of sequences, therefore four weblogo's was constructed, one for each sequence subset and one for the patient cohort which specifically looked at the chemotactic and neurotoxic domains of HIV-1 tat. The patient cohort showed similarity to both the subtype B and subtype C strains as it contained both cysteine and serine residues in position 31. The implications of this observations is not clear, but may be elucidated with further functional studies. Positions 24-51 and 31-61 are responsible for the chemotactic and neurotoxic properties of Tat respectively (Mishra et al., 2008). Mutations in this area would therefore most likely result in the augmentation of any biological activity. Single amino acid polymorphisms could be observed throughout the chemotactic and neurotoxic domains, although, sequence WebLogos also identified short amino acid sequences which were conserved. Tat is a strong chemo-attractant for macrophages, monocytes and dendritic cells. Amino acid residues in the cysteine rich domain, a conserved Ile39 and a SYXR motif at position 46–49 of Tat are all critical for chemokine receptor binding and signal transduction.

3.5. Conclusion

Comparison of subtype B, subtype C (Indian and South African) and the study cohort sequences revealed a high level of variation in amino acid residues. The frequency of point mutations in the subtype B strains was slightly higher than those identified in the subtype C strains. The implications of these novel mutations on the transactivation of *tat* remains unclear, therefore further analysis is necessary. The use of signature pattern analysis goes beyond simple nucleotide or amino acid distance relationships and could be useful for molecular epidemiologic investigations of HIV transmission, especially linked transmissions and questions of dual infection.

CHAPTER FOUR

Functional analysis of the *tat* gene and protein

4.1. Introduction

As discussed in previous chapters a high level of variability as well as areas of conservation was observed throughout the tat protein in the study cohort. When compared to subtype B and subtype C sequences, cohort sequences showed similarity to both. Signature pattern analysis identified amino acid residues which were clade specific with more variation occurring in subtype B than subtype C. However a total of 171 amino acids were identified as mutually occurring amongst all sequence subsets and the HAND cohort. Among these was the amino acids identified in the dicysteine tat motif, which was predominantly C30C31 or C30S31, with a small percentage of subtype B and subtype C sequences having variations thereof. According to Ranga et al (2004) the C30S31 mutation renders HIV-1 subtype C defective for monocyte chemotaxis *in vitro*. It has also been reported that subtype C is characterized by lower transactivation, lower levels of induced apoptosis and ultimately a lower incidence of HAND (Roa et al., 2009).

Study cohort sequences were identified with a 29% frequency of the C30C31 mutation and a 68% frequency of the C30S31 mutation. In addition a novel mutation, C30A31, was identified in the cohort with a frequency of 3%. In view of the association between the dicysteine motif and HAND, these mutations were selected to investigate the biological function of tat in cell culture *in vitro*. Briefly, Tat genes containing the above mentioned mutations were constructed and cloned into mammalian expression vectors which were transfected into 293T cells. This chapter will discuss the construction of tat genes and investigate the level of apoptosis being induced by each mutation. An outline of the experiments used to construct and express *tat* genes in cell culture is indicated in **Figure 4.1**. Briefly, patient sample JO206 was used to construct an optimized subtype C clone and 3 *tat* genes containing mutations identified in the dicysteine motif: C30C31, C30S31, and C30A31. The genes were transformed into bacterial cells and cloned into CMV expression vectors and transfected into 293T cells. Apoptosis levels were visualized and quantified with flow cytometry. The experimental procedures outlined in **Figure 4.1** will be discussed in more detail in the materials and methods sections.

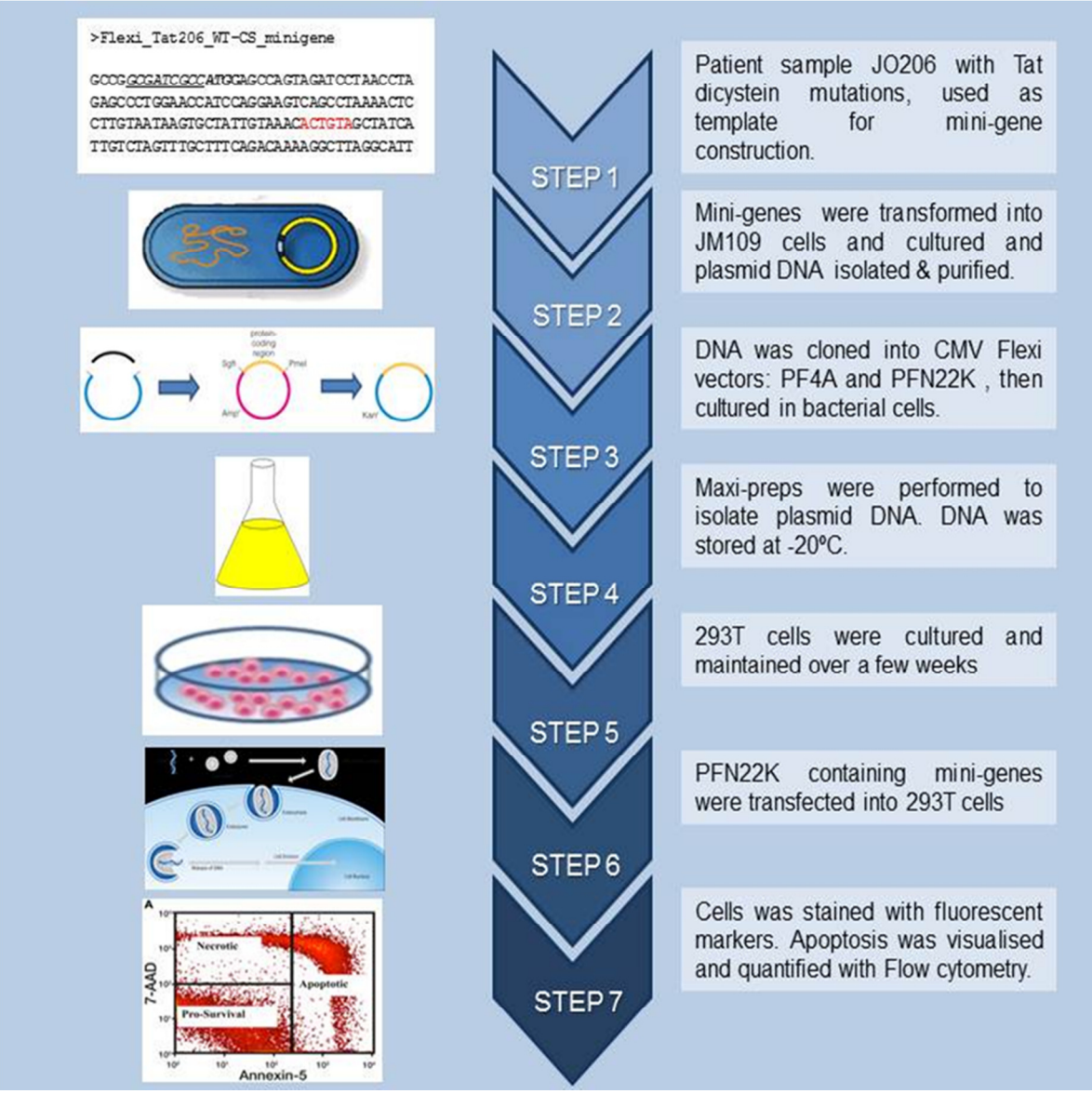

Figure 4.1 Overview of steps used to construct and express *tat* genes.

4.2. Materials and methods

4.2.1 Construction of *tat* mini-genes

Sequencing data was used to characterize the *tat* 1 and *tat* 2 genes of the study cohort as well as identify possible mutations present in the first *tat* exon. Two samples, JO206 and JO224 were identified as subtype C strains with the *tat* C30S31 polymorphism. These samples were closely associated to the consensus sequences of HIV-1 *tat* retrieved from the Los Alamos database (LANL HIV Database) however, JO206 was the only sample amplifiable for both the HIV-1 *tat* genes and it was therefore chosen as a template for the construction of the HIV-1 *tat* genes.

Four genes were constructed with the following polymorphisms in the *tat* motif: C30C31 (Subtype B wild-type mutation), C30S31 (Subtype C wild-type mutation) and C30A31 a novel mutation identified within the study cohort. An additional *tat* gene contained the subtype C wild-type mutation, however this gene had been codon optimized for expression in mammalian cells and would serve as a control. Each *tat* gene consisted of the complete *tat* 1 and *tat* 2 genes, start and stop codons for in frame expression of the *tat* protein, and restriction enzyme sites *Pme1* and *Sigel* for easy digestion and ligation into acceptor and donor vectors. Physical construction of the genes was performed by Integrated DNA Technologies (IDT) who also verified the ligation of *tat* gene into the pIDT-SMART-Amp vector with DNA sequencing.

4.2.2 Cloning of HIV-1 *tat* genes into pF4A cmvd1 flexi vector

Transformation of *tat* genes into JM109 cells

Tat genes were obtained from IDT at a concentration of 2µg/µl. The pIDT-SMART-Amp vectors were reconstituted in 10µl of nuclease free water to a working concentration of 200ng/µl. In order to maximize gene yield and plasmid volume a bacterial transformation was performed. Each reaction consisted of 50µl of competent JM109 cells; 200ng/µl of *tat* genes was added to the tubes containing the JM109 cells. Reactions were then allowed to incubate on ice for 20 minutes. Thereafter, reactions were transferred to a water bath at 42ºC and heat-shocked for 52 seconds. Reactions were incubated on ice for 2 minutes before adding 450µl of

SOC medium. Reactions were transferred to a shaking incubator for 90 minutes at 37ºC. Transformed cells were streaked in duplicate onto LB agar plates containing 100mg/ml ampicillin. Plates were inverted and incubated overnight at 37ºC. The pIDT–SMART-Amp vector contains the ampicillin resistance gene and makes colony selection possible. Colonies were picked and inoculated into LB broth containing ampicillin at a concentration of 100 mg/ml and incubated at 37ºC overnight. Plasmid DNA was isolated using the Pure yield midi prep kit (Promega, USA) according to manufacture specifications.

Digestion and ligation of mini-genes and pF4A cmvd1 flexi® vector

pF4A cmvd1 flexi® vector (Promega, USA) served as the first expression vector used to clone *tat* genes. pF4A cmvd1 flexi® vector is a mammalian expression vector constructed with cmv for both stable and transient gene expression of intrinsic proteins. Digestion of acceptor vector and pIDT–SMART-Amp vector plasmid DNA was performed concurrently using the restriction enzymes *Sgf* I and *Pme* I in 5X flexi® digest buffer and nuclease-free water. Both reactions were assembled according to the manufacturer's specifications and digested for 30 minutes at 37°C. The flexi® vector reaction was further heated to 65°C for 20 minutes to inactivate the restriction enzymes. Membrane binding solution (20µl) was added to the reaction containing the digested tat gene. Reactions were stored on ice. DNA concentrations of each reaction were determined using the nanodropTM ND-1000 system (Nanodrop Technologies Inc, USA). In a 0.2ml tube, 5µl of the pF4A cmvd1 flexi® vector acceptor vector (50ng) was mixed with 100ng *tat* gene DNA. To this 1µl (20U/µl) of T4 DNA ligase (Promega, USA) and 10µl of 2X flexi ligase buffer (Promega, USA) were added. The final volume of the reaction was brought to 20µl by adding nuclease-free water (Promega, USA) to the reaction and incubated at room temperature for 1 hour.

Transformation and bacterial culturing of recombinants

The pF4A cmvd1 flexi® vector also contains an ampicillin resistance gene which facilitates colony selection in *E. coli* grown on LB agar containing ampicillin. High-efficiency JM109 competent cells (Promega, USA) were thawed on ice and gently mixed before 50µl of the cells was transferred into a 14ml polypropylene round-

bottom tubes (Becton Dickinson, USA) containing 2µl ligated DNA. Tubes were incubated on ice for 20 minutes then heat-shocked for 45-50 seconds in Haake L water bath with Haake D1 heating circulator (LabX, USA) at 42℃. Tubes were immediately returned to ice for 2 minutes and 450µl of room-temperature SOC Medium (Sigma, USA) was added to each tube. Tubes were incubated for 1.5 hours at 37℃ with shaking **(**150 rpm). Each transformation (50-100µl) was plated onto duplicate plates containing 100µg/ml ampicillin and incubated overnight at 37℃. After 24 hours, colonies were picked and inoculated into 250ml LB broth containing ampicillin at a concentration of 100 mg/ml and incubated at 37ºC overnight.

Bacterial cell cultures were grown for approximately 16 hours and purified using the Pure-Yield Plasmid Midi-prep System from Promega (Promega, USA). Bacterial cell suspension was centrifuged at 5000 X *g* for 10 minutes. The supernatant was discarded and replaced with 6ml of cell re-suspension solution and inverted 3-5 times. Cells were incubated for 3 minutes at room temperature before adding 6ml of cell lysis solution and inverting 5-10 times. Solution was inactivated by 10ml neutralization solution inverted 5-10 times and centrifuging at 15000 X *g* for 15 minutes. The cleared lysate was decanted into a PureYield Clearing Column and the filtered DNA bound to the PureYield Binding Column. The binding column was washed once with 5ml of endotoxin solution and 20ml column wash solution. The membrane was dried by applying a vacuum for 1 minute, purified DNA was eluted in 400µl of nuclease free water. Plasmid DNA concentrations was determined with the Nanodrop TM ND-1000 system.

4.2.3 Cloning of HIV-1 *tat* genes into pFN22K cmvd1 flexi vector

Digestion and ligation of cmvd1 flexi vectors

pFN22K cmvd1 flexi® vector (Promega, USA) was used as the second expression vector for cloning of the *tat* genes. The 5X flexi® digest buffer, pFN22K cmvd1 flexi® vector and nuclease-free water were thawed and stored on ice. The reaction was assembled according to manufacturer's specifications with 100ng of both pF4A and pFN22K cmvd1 flexi vector DNA digested with restriction enzymes *Sgf* I and *Pme* I for 30 minutes at 37℃. The final volume of the reaction was brought to 20µl by adding nuclease free water (Promega, USA) to the reaction. The reaction was

further heated to 65°C for 20 minutes to inactivate the restriction enzymes and stored on ice until ligated. In a 0.2ml tube 10µl of the digested pFN22K and pF4A cmvd1 flexi® vector mix (100ng) was pipetted along with 10µl of 2X Flexi Ligase Buffer and 1µl of T4 DNA ligase (20U/µl). Nuclease free water was added to the reaction for a final volume of 21µl. The reaction was incubated at room temperature for 1 hour.

Transformation and culturing of pFN22K cmvd1 flexi vector with *tat* gene

High-efficiency JM109 competent cells (Promega, USA) were thawed on ice and gently mixed before 50µl of the cells was transferred into a 14ml polypropylene round-bottom tubes (Becton Dickinson, USA) containing 2µl ligated DNA. Tubes were incubated on ice for 20 minutes then heat-shocked for 45-50 seconds in Haake L water bath with Haake D1 heating circulator (LabX, USA) at 42°C. Tubes were immediately returned to ice for 2 minutes and 450µl of room-temperature SOC medium (Sigma, USA) was added to each tube. Tubes were incubated for 1.5 hours at 37°C with shaking **(**150 rpm). Each transformation (50-100µl) was plated onto duplicate plates containing 25 mg/ml kanamycin and incubated overnight at 37°C. After 24 hours, colonies were picked and inoculated into 250ml LB broth containing kanamycin at a concentration of 25 mg/ml and incubated at 37ºC overnight. The PFN22K cmvd1 flexi® vector contains a kanamycin resistance gene which facilitates colony selection in *E. coli* grown on LB agar containing kanamycin.

Bacterial cell cultures were grown for approximately 16 hours. Plasmid DNA was then isolated and purified using the Pure-Yield Plasmid Midi-prep System from Promega (Promega, USA). Bacterial cell suspension was centrifuged at 5000 X *g* for 10 minutes. The supernatant was discarded and replaced with 6ml of cell re-suspension solution and inverted 3-5 times. Cells were incubated for 3 minutes at room temperature before adding 6ml of cell lysis solution and inverting 5-10 times. Solution was inactivated by 10ml neutralization solution inverted 5-10 times and centrifuging at 15000 X *g* for 15 minutes. The cleared lysate was decanted into a PureYield Clearing Column and the filtered DNA bound to the PureYield Binding Column. The binding column was washed once with 5ml of endotoxin solution and 20ml Column Wash solution. The membrane was dried by applying a vacuum for 1

minute, purified DNA was eluted in 400µl of nuclease free water. Plasmid DNA concentrations was determined with the Nanodrop TM ND-1000 system.

Screening of colonies for *tat* inserts

Tat gene inserts in both the pFA4A and pFN22K CMVd1 Flexi® Vectors were verified with colony PCR using the primers Tat F 5´-AGAGAAGAAAGCTTAATGGAGCCAGTAGATC-3´ and Tat R 5´-ACTTACAATTGTGATATAAGATTTTGATG-3´. Reagent concentrations and cycle parameters have been represented in **Tables 4.1 and 4.2.** Reactions were inoculated with colonies also inoculated in LB broth for overnight cultures The GeneAmp® 9700 PCR system (Applied Biosystems, USA) was utilized for amplification of inserts with cycle parameters described in **Table 4.2.** The initial denaturing step lyses bacterial cell walls and releases plasmid DNA which could now be amplified and visualized with ethidium bromide and gel electrophoresis.

Table 4.1 Mastermix composition of colony PCR

Reagent	Final Concentration	Volume µl
dNTPs (10mM each)	0.2mM	1
Tat F Primer (40pmol/µl)	0.8pmol/µl	1
Tat R Primer (40pmol/µl)	0.8pmol/µl	1
25 mM $MgCl_2$	1.5mM	3
5X GoTaq Flexi buffer	1X	10
5U/µl Taq DNA	1.1U/µl	0.75
Water	-	32.25
Total volume	50µl	50µl

Table 4.2 Cycle parameters of colony PCR

	Temperature	Time	Cycles
Initial Denaturation	94 ºC	2 minutes	1 X
Denaturation	94 ºC	30 seconds	40 X
Annealing	55 ºC	30 seconds	
Primer extension	68 ºC	3 minutes	
Final extension	68 ºC	10 minutes	1X
Hold	4 ºC	∞	

4.2.4 Cell culture

Cell line

HEK 293 cells, an adherent cell line, were provided by Dr Craig Kinnear from the Genetics Department of Stellenbosch University and were used to investigate the apoptotic effects of constructed mini-genes in cell culture. HEK 293 cells are commonly referred to as 293T cells and are derived from human embryonic kidney cells. This cell line has been described as being highly transfectable and supports elevated levels of protein expression.

Maintenance of Cell line

The 293T cell line was cultured in a LabotecTM Biological Safety Cabinet Class II (Labotec, Germany) to protect the cell line from contamination. Cells were removed from liquid nitrogen and thawed with 10ml of DMEM (Lonza, Switzerland) supplemented with 10% foetal bovine serum (FBS) (Lonza, Switzerland) and 0.1mg/ml streptomycin (Lonza, Switzerland). Cells were centrifuged twice at 150 X *g* for 5 minutes to remove all DMSO from cell suspension. Cells were seeded into 25cm^2 Cell Culture Flasks (Corning, USA) and incubated at 37ºC supplemented with 5% CO_2 in NUAIRETM US Autoflow CO_2 Water-Jacketed Incubator (Nuaire, USA). After 48 hours culture media was replaced with fresh DMEM media.

Regular sub-culturing was performed to maintain the growth of the cell line. Once cells had reached 80% confluence they would be sub-cultured into new culture flasks. Old culture medium would be discarded and cells would be washed with 10ml of 1 XPBS (Lonza, Switzerland) to remove excess FBS from culture flask which may inhibit the effects of trypsin. Trypsin /versene (Trypsin/EDTA), a combined reagent (Lonza, Switzerland) (1.5ml) was added to the monolayer and gently shaken remove cells from the culture vessel surface into suspension solution. Cells were re-suspended in 5ml of culture medium and transferred to a sterile 15ml High-Clarity Polypropylene Conical Tube (Becton-Dickenson, USA). Cell suspension was centrifuged at 171 X *g* for 5 minutes. A second wash step was performed to ensure removal of trypsin. Cell pellet was re-suspended in 5ml culture medium.

In order to determine the correct seeding density for sub-culturing, the viable and total cell count was determined. Cell suspension was vortexed for even distribution; 50µl of the cell suspension was removed and mixed with an equal volume of 0.4% trypan blue stain (Invitrogen, USA). Ten microliters of the mixture was transferred to a Neubauer Improved Bright-Line (Marienfeld Superior, Germany) cell counter. Healthy (unstained) and dead/dying (stained) cells were counted using an Eclipse E200 Microscope (Nikon, USA). The cell count together with the formula below was used to determine the percentage viable cells per millilitre. Total cells counted / number of squares × 10^4 × dilution factor = Number of viable cells/ml. Cells were diluted with complete DMEM to obtain a cell concentration of 1X10^6cells/ml and seeded into a new 75 cm^2 cell culture flasks (Corning, USA) containing 20 ml of complete DMEM medium. Cells were then incubated in a Nuaire^{TM} US Autoflow CO_2 water-jacketed incubator (Nuaire, Minnesota, USA). A second cell concentration of 1X10^6cells/ml was used for the storage of excess 293T cells.

Mycoplasma screening

Cultured cell lines are repeatedly passaged and can therefore be easily contaminated by mycoplasma. Mycoplasma contamination is a major problem in biological research and can alter the metabolism of cells and ultimately influence the outcome of generated data. Cells were therefore screened for mycoplasma contamination using RIDASCREEN® Mycoplasma IFA (R-Biopharm AG, Darmstadt, Germany). A total of 25 000 cells were seeded into trac bottles containing a coated

glass microscope slide and 2ml culture medium and incubated for two days. Trac bottles were rinsed with 1X PBS and cells fixed with cold acetone for 5 minutes. The coated glass microscope slide was removed and attached to a glass slide with glue; the sample was then air-dried at room temperature. One drop of the anti-mycoplasma-conjugate was added to the fixed cells and incubated for 20 minutes. After the incubation the slide was rinsed with phosphate buffered saline and washed twice for a total of 2 minutes in a bath of 1XPBS. The slides were incubated at room temperature to dry. One drop of anti-mycoplasma-conjugate was added to the center of the slide and incubated for another 20 minutes. After the incubation the wash step was repeated and the slide was air-dried at room temperature. One drop of mounting fluid was added in the center of each well and a coverslip was placed on it. The slides were observed with the Olympus BX60 fluorescent microscope (Wirsam Scientific, RSA) and screened for the presence of green fluorescent inclusion bodies within cells.

4.2.5 Transfection and optimization of apoptosis assay

Transfection control

In order to ensure that the detected level of apoptosis in cell culture was a result of the mini-genes only, a transfection control (pFN22K CMVd1 Flexi® Vector) was used which had been digested with Flexi® Enzyme Blend (*Sgf* I and *Pme* I) (Promega, USA) to remove the barnase gene. Since the vector cannot ligate upon itself, oligonucleotides, Flexi-strand-F: 5´-CGCGTAAGGGTAGGTTT-3´ and Flexi-strand-R: 5´-AAACCTACCCTTACGCGAT-3´ were incorporated into the vector. The vector was then transformed into JM109 competent cells as described by Dr Bizahn Romani (PhD thesis, Stellenbosch University, 2011). Tat Clones and transfection control were transfected using the Extreme gene 9 transfection reagent (Roche, USA) according to manufacturer specifications in a 3:1 ration of 3µl of transfection reagent and 1µg/µl Tat clone DNA.

Apoptosis optimization

Flow cytometry was used to assess the apoptosis-inducing efficacy of the constructed *tat* mini-genes in cell culture. Intrinsic and extrinsic stimuli can induce cell death. During apoptosis, cells can be in one of three stages of cell death: early

apoptosis, late apoptosis or necrosis. Flow cytometry parameters for the detection of Annexin V-FITC (BD Biosciences, USA) and 7-AAD (Sigma, USA) fluorescent dyes were optimized to identify 293T cells undergoing cell death. Furthermore the use of FITC with 7-AAD excludes any spectral overlap and concerns regarding color compensation as they are detected in channels 1 and 3 of the FACS Calibur (BD, USA). Annexin V is a phospholipid-binding protein with a high affinity for phospholipid phosphatidylserine (PS). During early apoptosis the PS molecule translocates from the inner to the outer leaflet of the plasma membrane, thereby exposing PS to annexin V FITC. During late phase apoptosis the PS molecule remains translocated, the plasma membrane is degraded and permealised allowing 7-AAD to enter the cell and bind to DNA in the nucleus. Cells staining positive for 7-AAD only, were undergoing necrosis. In addition camptothecin (Sigma, USA), a strong apoptosis inducer, was used to induce apoptosis (positive control) for optimization of flow cytometry (Hsiang *et al.*, 1985; Wall *et al.*, 1966).

In a 6 well plate 250 000 cells were seeded and incubated for 18 hours at 37 ºC. Once cells had reached 70-80% confluence they were stimulated with 5µM camptothecin for 4 hours. Thereafter cells were washed with warm 1XPBS and trypsinized with 2mM EDTA/1XPBS for 1 minute. Cells were washed twice with 1XPBS and centrifuged at 150 X *g* for 5 min according to manufacturer's instructions (FITC Annexin V, BD Pharmingen). Viable cells were counted prior to antibody staining. A total of $1X10^6$ cells were resuspended 1ml 1× Binding Buffer, 100µl of the solution ($1X10^5$ cells) was transferred to a 5 ml culture tube and stained with 5µl Annexin V FITC and 5µl 7-AAD and incubated for 15 minutes in the dark. Cells were washed in 450µl binding buffer solution and centrifuged at 150 X *g* for 5 minutes to remove unbound antibody. Cells were resuspended in 450µl binding buffer solution before visualizing apoptosis with flow cytometry.

4.2.6 HIV-1 Tat induced apoptosis and flow cytometry assays

Transfection of 293T cells with *tat* mini-genes

Cells intended for transfection were seeded in 6 well plates at a concentration of 250 000 cells and incubated at 37ºC supplemented with 5% CO_2 in NUAIRETM US

Autoflow CO_2 Water-Jacketed Incubator (Nuaire, USA) overnight. Once cells had reached approximately 70% confluence old tissue culture medium was replaced with fresh medium. A transfection complex consisting of 100µl DMEM, 3µl transfection media and 1µg/µl DNA was prepared and allowed to incubate for 15 minutes before being added to 293T cells in a drop-wise manner. Cells were incubated at 37ºC supplemented with 5% CO_2 in NUAIRETM US Autoflow CO_2 Water-Jacketed Incubator (Nuaire, USA) for 18 hours.

Antibody staining of transfected 293T Cells

Cell populations were trypsinized with 2mM EDTA/1XPBS for 1 minute. Cells were centrifuged at 150 X *g* for 5 minutes and the supernatant discarded. Cells were then washed with 1X PBS and centrifuged at 150 X *g* for 5 minutes and the supernatant discarded, the wash step was repeated once more before $1X10^5$ cells were re-suspended in 450µl binding buffer. Cells were stained with 5µl of Annexin V and 5µl 7-AAD and incubated in the dark for 15 minutes. Cells were washed as previously described and then resuspended in 450µl binding buffer. Samples, negative control tubes, consisting of a pSV-β-Galactosidase vector control and transfection media only control tubes and one positive control tube containing camptothecin were analyzed on a four color FACS Calibur flow cytometer (Becton Dickson, USA).

Identification and gating of apoptotic cell populations

Apoptotic and non-apoptotic cells were differentiated from each other with the use of flow cytometry using the FACS Calibur (Becton Dickenson (BD) Bioscience, USA) and CellQuest Pro (Becton Dickenson (BD) Bioscience, USA). A maximum of 10 000 – 20 000 cell events were acquired. The cells were gated based of forward scatter vs. side scatter (to exclude debris and or cell clumps) and then further divided into four cell populations using a quadstat dot-plot comparing the two fluorescent dyes as indicated in **Figure 4.6.** Cell populations consisted of non-apoptotic cells which were negative for both antibodies, early apoptotic cells which were negative for 7-AAD and positive for Annexin V FITC, late apoptotic cells were positive for both Annexin V FITC and 7-AAD while necrotic cells were positive for 7-AAD. Quadrant statistics were generated after identifying observed apoptotic cells and used in a camptothecin

optimized assay to determine the percentage of cells in each described apoptotic stage. Mean percentage values were used to determine the level of apoptosis induced by each clone, and were compared and analyzed to determine any statistical significance. Statistical analysis was performed in graphpad and a two way analysis of variance (ANOVA) for repeated measures was performed to compare differences in time points. Samples with a p value ≤ 0.05 were deemed significant. Instrument settings and color compensation was performed using stained beads (BD Biosciences, USA). Manual compensation, to obtain optimal visualization of cell populations, was performed using 293T cells which were either untreated or induced into apoptosis with camptothecin. Color compensation was not a concern as the two fluorescent markers used to stain cells did not have any spectral overlap. Unstained cells were used to ensure that non-specific or auto fluorescent signal detection did not occur. These settings were used for subsequent apoptosis assays.

4.3. Results

4.3.1 Construction of HIV-1 *tat* genes

Construction of mini-genes were performed by IDT and sequenced to ensure that the *tat* gene sequence was in frame for proper protein expression. Each sequence was cloned into a pIDTSMART vector containing M13 forward and reverse sites, an ampicillin resistant gene and a site of origin. The vector was 2434bp in size and 1.503g/mol. In addition, the vector also contained sites for restriction enzyme Pme1 and Sgf1 for easy transfer between CMV flexi vectors (Promega, USA). The sequences for each constructed tat gene are indicated in **Figure 4.2** below:

>Flexi_Tat206_WT-CS_minigene
GCCG*GCGATCGCC**ATG*GAGCCAGTAGATCCTAACCTAGAGCCCTGGAACCATCCAGGAAGTCAGCCTAAAACTCC
TTGTAATAAGTGCTATTGTAAACACTGTAGCTATCATTGTCTAGTTTGCTTTCAGACAAAAGGCTTAGGCATTTC
CTATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGCAGTGAGGATCATCAAAATCTTATATCAAA
GCAGCCCTTACCCCAAACCCGAGGGGACCCGACAGGCTCGGAAGAATCGAAGAAGAAGGTGGAGAGCAAGACAGA
GACAGATCCATTCGATGTT***TAA***ACACGG

>Flexi_Tat206_CC_minigene
GCCG*GCGATCGCC**ATG*GAGCCAGTAGATCCTAACCTAGAGCCCTGGAACCATCCAGGAAGTCAGCCTAAAACTCC
TTGTAATAAGTGCTATTGTAAACACTGTTGCTATCATTGTCTAGTTTGCTTTCAGACAAAAGGCTTAGGCATTTC
CTATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGCAGTGAGGATCATCAAAATCTTATATCAAA
GCAGCCCTTACCCCAAACCCGAGGGGACCCGACAGGCTCGGAAGAATCGAAGAAGAAGGTGGAGAGCAAGACAGA
GACAGATCCATTCGATGTT***TAA***ACACGG

>Flexi_Tat206_CA_minigene
GCCG*GCGATCGCC**ATG*GAGCCAGTAGATCCTAACCTAGAGCCCTGGAACCATCCAGGAAGTCAGCCTAAAACTCC
TTGTAATAAGTGCTATTGTAAACACTGTGCCTATCATTGTCTAGTTTGCTTTCAGACAAAAGGCTTAGGCATTTC
CTATGGCAGGAAGAAGCGGAGACAGCGACGAAGCGCTCCTCCAAGCAGTGAGGATCATCAAAATCTTATATCAAA
GCAGCCCTTACCCCAAACCCGAGGGGACCCGACAGGCTCGGAAGAATCGAAGAAGAAGGTGGAGAGCAAGACAGA
GACAGATCCATTCGATGTT***TAA***ACACGG

>Flexi_Tat206_OPT_minigene
GCCG*GCGATCGCC**ATG*GAGCCCGTGGACCCCAACCTGGAGCCCTGGAACCACCCCGGCAGCCAGCCCAAGACCCC
CTGCAACAAGTGCTACTGCAAGCACTGCAGCTACCACTGCCTGGTGTGCTTCCAGACCAAGGGCCTGGGCATCAG
CTACGGCCGCAAGAAGCGCCGCCAGCGCCGCAGCGCCCCCCCCAGCAGCGAGGACCACCAGAACCTGATCAGCAA
GCAGCCCCTGCCCCAGACCCGCGGCGACCCCACCGGCAGCGAGGAGAGCAAGAAGAAGGTGGAGAGCAAGACCGA
GACCGACCCCTTCGACgtt***TAA***acacgg

Figure 4.2: Nucleotide Sequence of constructed Mini-genes. Four tat genes were constructed each with a different mutation in the 30/31 positions of tat. The presence of start and stop codons, restriction enzymes sites and the 30/31 positions are indicated

4.3.2 Screening of *tat* genes in donor vector

Tat gene transfer between donor (pF4A CMVd1 Flexi® Vector) and acceptor vectors (pFN22K CMVd1 Flexi®Vectors) was verified by the amplification of the *tat* gene. Four colonies per mini-gene were screened for the presence of *tat* inserts. Positive products were visualized with gel electrophoresis and ethidium bromide. These results indicated that the cloning of the *tat* mini-genes was successfully completed in both the donor and acceptor vector.

4.3.3 Detection of possible mycoplasma contamination

Mycoplasma infection has been shown to alter the normal functioning of cell lines such as growth, viability, metabolism and biological properties and can therefore

become a serious problem once cells have been contaminated. It is especially difficult to identify as it cannot be detected with normal light microscopy. It is therefore important to screen cell lines for mycoplasma infection on a regular basis. Cell cultures were screened with RIDASCREEN® Mycoplasma IFA, an assay which utilizes monoclonal antibodies for the detection of a broad spectrum of Mycoplasma species which include *Acholeplasma laidlawii*, *Mycoplasma hyorhinis*, *M. arginini*, *M. orale*, *M. fermentans* and *M. salivarium*, which account for more than 96% of cell culture infections (Blazek *et al.*, 1990; Kamla *et al.*, 1992). Initial screening of cell cultures with RIDASCREEN® Mycoplasma IFA appeared suspicious however, it was not clear if the culture was indeed infected with mycoplasma or perhaps another microorganism. The possible contamination of the cell culture resulted in the entire cell line to be discarded as well as any media being used at that time. All surfaces used for the culturing and incubation of cells were disinfected with Biocide, Alcohol and Contrad, fresh media and a new vial of cells was also cultured to ensure the cell line was free of any contamination. The new cell culture was again screened for infection with RIDASCREEN® Mycoplasma IFA as indicated in **Figure 4.3**. The immunofluorescence assay did not show any signs of contamination and could therefore now be used for transfection and other experiments. No other incidents of contamination occurred with the growth of new 293T cells and the use of new growth media.

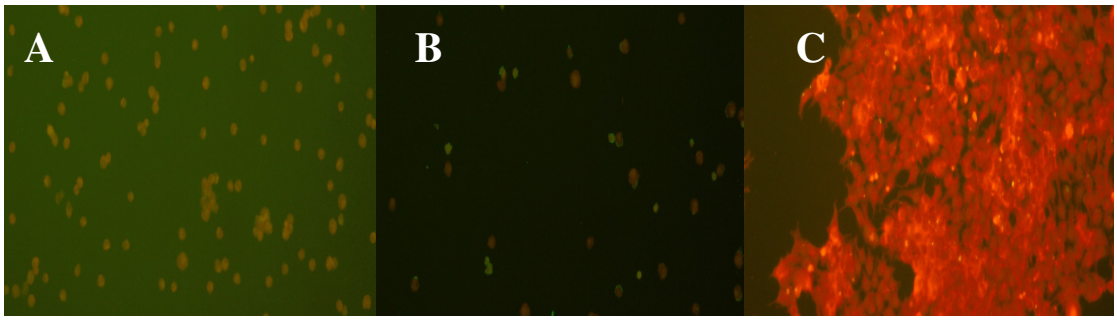

Figure 4.3: Immunofluorescence staining for the presence of mycoplasma infection. (A) Negative mycoplasma control, **(B)** positive mycoplasma control, cells stained with green inclusions indicates the presence of mycoplasmas, **(C)** 293T cell cultures intended for transfection, were screened for the presence of mycoplasmas with immunofluorescence staining. No contamination by mycoplasmas was detected.

4.3.4 Optimization of apoptosis assay

Camptothecin, a cytotoxic quinolone alkaloid, is capable of inhibiting DNA enzyme topoisomerase I (*topo* I) and is commonly used in apoptotic assays. Camptothecin was used here to optimize flow cytometry detection using two concentrations (2.5µM and 5µM) to induce apoptosis in the 293T cells at three time periods, 4hours, 6hours and 24hours. Optimum apoptosis conditions was observed with 5µM camptothecin incubated for 4 hours as longer incubation periods showed high levels of necrosis or complete decimation of 293T cells. **Figure 4.4** illustrates the forward vs. side scatter dot plot used to exclude debris and gate on live and apoptotic cells.

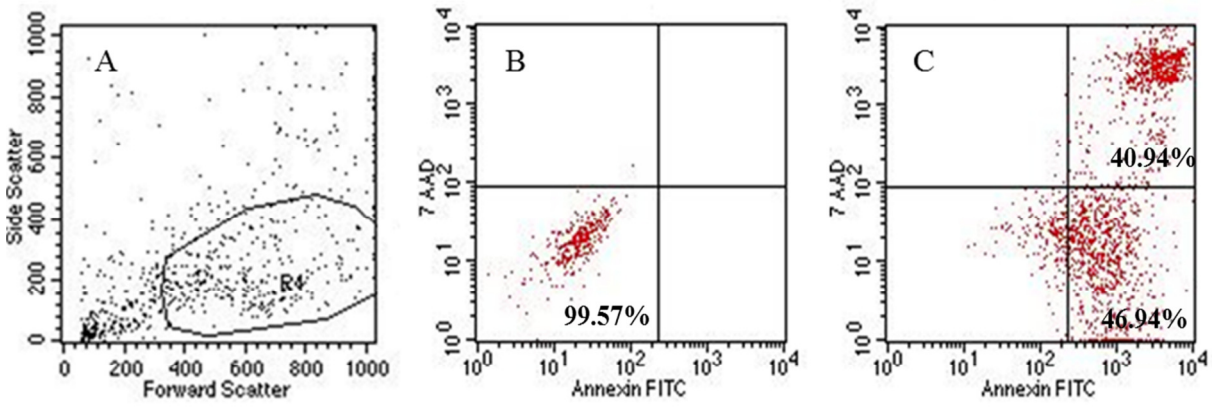

Figure 4.4: Gating strategy to identify non-apoptotic and apoptotic cell populations. (A) Cell of interest are isolated using a gate on a forward and side scatter plot **(B)** Untreated cells are found in the lower left quadrant of a 7-AAD vs. Annexin V FITC scatter plot **(C)** while early and late apoptotic cells could be identified in the lower right and upper right quadrants respectively.

4.3.5 Apoptosis induction of 293T cells by *tat* mini-genes

Apoptosis analysis was performed on the flow cytometer using CellQuest Pro (BD, USA) software. A scatter plot was created and cells were gated to exclude doublets, a second plot was generated from the gated cells as indicated in **Figure 4.5.** Cells that were positive for the Annexin V marker were identified as early apoptotic cells, whereas cells staining positive for both 7-AAD and the Annexin V markers were labeled as late apoptotic cells. Cells which were negative for both fluorescent dyes were considered as non-apoptotic while cells positive for 7-AAD only were considered necrotic.

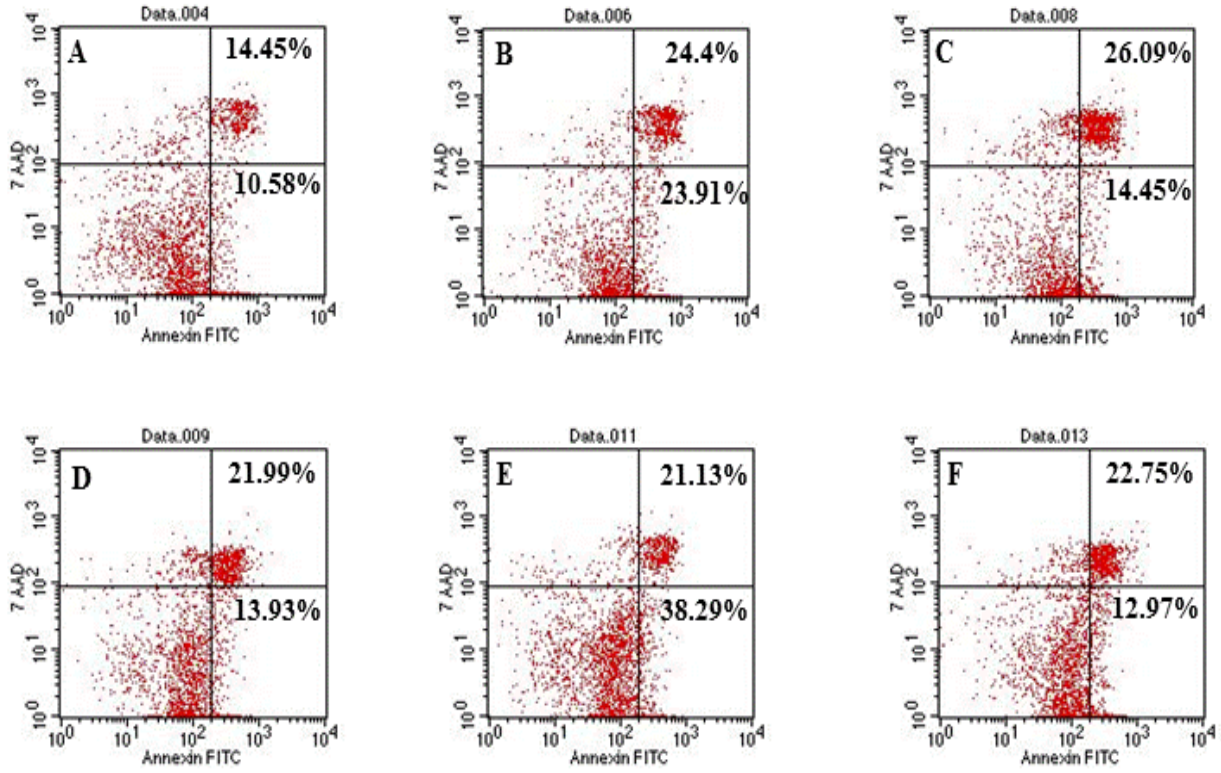

Figure 4.5 Percentage of apoptotic 293T cells by Annexin V and 7-AAD. Scatter plots represents apoptotic results for **(A)** transfection media control, **(B)** vector only control, **(C)** C30C31 clone, **(D)** C30S31 clone, **(E)** Optimized clone and **(F)** C30A31 clone. Different cell populations could be identified in each quadrant of the scatter plots: Non-apoptotic cells in the lower left quadrant, early apoptotic cells in the lower right quadrant, late apoptotic cells in the upper right quadrant and necrotic cells in the upper left quadrant. Percentages of cells undergoing apoptosis have been included in quadrants of interest.

Assays were performed in triplicate to determine an apoptotic trend amongst the Tat clones however the level of apoptosis from one experiment to the other was not reproducible for all cell populations. Differences in cell viability occurred despite cell cultures and Tat clones being treated the same and could possibly have influenced apoptosis assays. In addition varying degrees of necrosis was also detected as evident by scatter plots (A-F) in **Figure 4.5**. Cell viability for each clone and control has been shown in **Table 4.1** for each apoptosis assay.

Table 4.1: Detected cell viability in control and *tat* clones

	Cell Viability (%)					
	Transfection media	**Vector**	**Optimized clone**	**C30C31 clone**	**C30S31 clone**	**C30A31 clone**
Experiment 1	57	53	78	75	56	87
Experiment 2	63	55.36	68.75	72.07	76.47	68
Experiment 3	50	66.67	54.55	60.47	71.43	56
Mean value	56.67	58.34	67.1	69.18	67.97	70.33
Standard deviation	6.51	7.31	11.81	7.68	10.67	15.63

The levels of apoptosis were determined as a percentage of apoptotic cells in the entire cell population. The total mean percentages of apoptosis induced by Tat clones and controls have been indicated below in **Table 4.2.** Overall, all Tat clones displayed total mean apoptotic levels greater than 20% except the C30S31 clone which was identified with the lowest level of total mean apoptotic events at 17%. The optimized Tat clone was identified as having the greatest level of apoptotic events with a total mean apoptotic percentage of 37.45%.

Table 4.2 Total mean percentage of apoptosis induced by Tat clones and controls

Tat clones and Controls	**Total mean percentage of apoptosis detected (%)**	**Clade specific apoptosis (%)**
Transfection media (neg)	16.00 (± 9.65 SD)	-
Vector (neg)	22.34 (± 23.22 SD)	-
CA clone	26.34 (± 13.49 SD)	4
CS clone	17.00 (± 17.69 SD)	-
CC clone	28.36 (± 10.83 SD)	6.02
Optimized clone	37.45 (± 19.08 SD)	15.11

Analysis of early and late apoptosis data identified a trend amongst Tat clones with most cells being characterized as being late apoptotic as illustrated in figure 4.6. The optimized Tat clone was the only exception to this trend and displayed the highest

level of early apoptotic events with a mean percentage of 20.34%. The C30S31 Tat clone was identified with the lowest mean percentage for both early (5.87%) and late (11.13%) apoptotic events while the C30C31 was identified as having the highest mean percentage for late (18.96%) apoptotic events. Surprisingly the Tat clone with the C30A31 mutation, displayed a similar pattern of apoptosis as the Tat clone containing the C30C31 mutation. The Tat clone with the C30A31 mutation had a mean percentage of 7.92% for early apoptosis and 18.42% for late apoptosis.

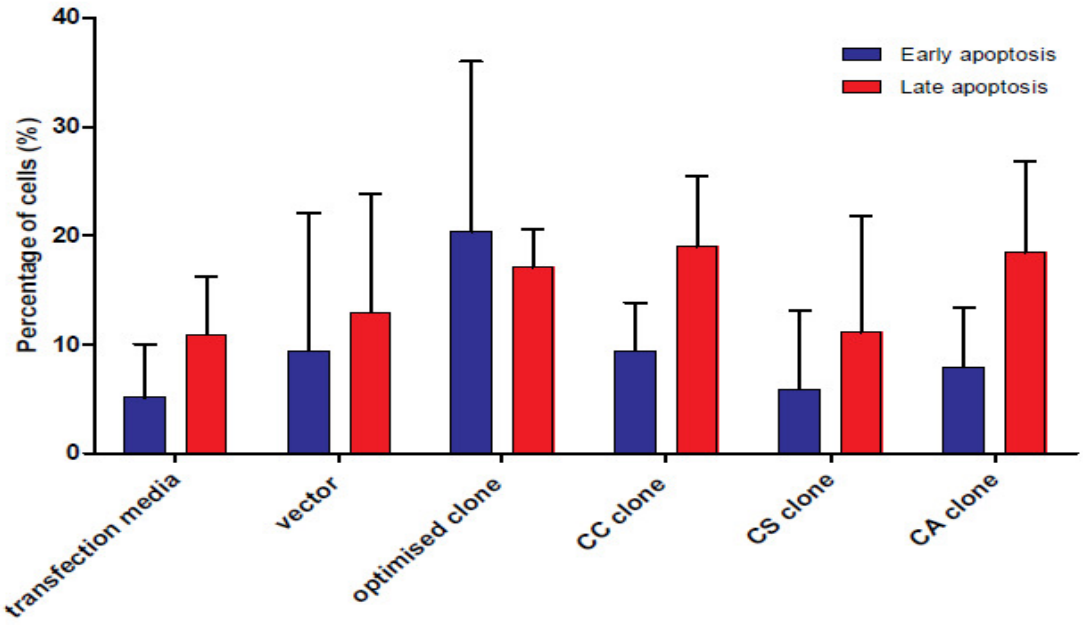

Figure 4.6 Percentage of cells undergoing early and late apoptosis. The figure illustrates, on the Y-axis, the percentage of cells undergoing early or late apoptosis with controls and samples being indicated on the X-axis. The error bars indicate the standard deviations in the reported measurements done in triplicate. The transfection media and empty vector were used as negative controls. Four samples were analyzed consisting of an optimized clone, C30C31 clone, C30S31 clone and a C30A31 clone.

A *t*-test for dependant samples and a 2-way ANOVA were performed to determine the statistical significance of the level of apoptosis being induced by each T*at* clone as well as the levels of early and late apoptotic events. However, no statistical significance could be observed for both tests. The 2-way ANOVA analysis between

early and late apoptosis generated a *p*-value of 0.83, while the *p*-values for the interaction between Tat clones were all $p>0.05$.

4.4. Discussion

Transfection and Apoptosis induction of 293T cells by *tat* mini-genes

HIV-1 Tat has been associated with the pathogenesis of neuro-AIDS and has been identified as a highly toxic protein capable of inducing apoptosis in astrocyte and microglial cell cultures (Mishra et al., 2010). Tat is most abundant in its 101 amino acids form, which is characteristic of subtype C infections. However , truncated forms composed of 72 and 86 amino acids have been more commonly used in HIV-1 Tat studies, especially Tat HXB2 (86), a laboratory-passaged subtype B strain (Campbell et al., 2005; Li et al., 2012). Here, we investigate the apoptotic effects of full length Tat proteins, containing subtype B (C30C31) and subtype C (C30S31, C30A31) mutations expressed in 293T cell culture by CMV mammalian expression vectors.

Analysis of cells undergoing apoptosis were divided into one of three categories, early apoptotic events: cells staining positive for Annexin V, late apoptotic events: cells dual positive for Annexin V and 7-AAD and total apoptotic events: the sum of both early and late apoptotic events. Overall, the optimized clone generated the highest level of total apoptosis of 37.45%. This was anticipated, as this clone had been codon optimized for protein expression in mammalian cells. C30C31, C30A31, and C30S31 showed total mean apoptosis levels of 28.36%, 26.34% and 17.00% respectively. The low levels of apoptosis observed in C30S31 correlated to previous reports from another study by Mishra et al (2010) which showed that the C31S mutation resulted in lower levels of apoptosis *in vitro*. However it was not expected that the level of apoptosis observed with the C30S31 mutation would be lower than the vector only control which had a mean apoptosis level of 22.34%. Yet, C30S31 apoptosis induction was still higher than the transfection media control (16%), though not by much. This would suggest that the vector alone was capable of inducing apoptosis. However, this observation could have been confounded by the variability in cell viability observed during the experiments. Cell viability was reasonably high

and ranged between 67.1% and 70.33% for Tat clones while cell viability for transfection media and vector only controls was 56.67% and 58.34% respectively. Assessing the apoptotic capabilities of Tat clones has therefore been more challenging as pseudo apoptosis levels observed in controls could have been associated with cell viability. To overcome this problem future study should make use of a transfection marker, such as Halo tag, to isolate cells which have been transfected and are undergoing apoptosis.

Nevertheless, the two clones C30C31 and C30A31 were still able to induce apoptosis above both control levels. Furthermore, the novel mutation C30A31 produced early and late apoptosis levels similar to the apoptotic levels observed with the clone with the C30C31 mutation. For the clone with the C30A31 mutation, a mean apoptotic level of 7.92% was observed for early apoptosis and 18.42% was observed for late apoptosis, while the clone with the C30C31 mutation induced apoptotic levels of 9.39% for early apoptosis and 18.96% for late apoptosis. It is unclear if these are typical apoptosis levels for C30A31, as this mutation has previously not been described in the literature. Nonetheless, the levels of apoptosis induced by clones with the C30C31 mutations were comparable to those reported in a study conducted by Campbell et al. (2005), which compared apoptosis induction of full length Tat to that of the HXB2(86) form.

4.5. Conclusion

In summary, the findings of these experiments showed that Tat clones were capable of inducing apoptosis. The apoptosis induced by the Tat clone containing the C30C31 mutation was however more robust than that of the Tat clone with the C30S31 mutation. Therefore demonstrating that the naturally occurring subtype C mutation should be regarded as defective. However the C30A31 mutation has also demonstrated that subtype C is capable of apoptotic potency similar to that found in subtype B and highlights the importance of protein functionality of different HIV-1 subtypes.

CHAPTER FIVE

Overall Discussion and Conclusion

Table of contents	Page

CHAPTER FIVE

Overall Discussion and Conclusion

5.1 Discussion

In recent years the prevalence of HIV-1 in South Africa has resulted in a gradual plateau; however growing evidence has shown that many infected individuals also develop many neurological deficits which include encephalitis, cognitive slowing and HAND.

There is currently little information available regarding HAND cases especially in South Africa as many genetic and functional studies conducted on HAND patients have been focused on either subtype B or Indian subtype C strains. Studies have identified a multitude of pathways as well as genetic and cellular factors which contribute to the progression of HAND which include HIV-1 intrinsic and extrinsic apoptotic pathways, proteins such as HIV-1 Tat and chemokine's and cytokines such as Interferon α and Fas ligand. Furthermore factors such as subtype specificity have also been implicated as a possible contributing factor of HAND pathogenesis. This idea has been strengthened by the fact that most HAND cases occurred in individuals infected with subtype B strains than subtype C strains. However these results may be biased as subtype C is most prevalent in developing countries found in Africa and Asia where HAND prevalence may be poorly documented.

Nonetheless, it cannot be denied that subtype B and subtype C HIV strains are genetically distinct. This is most evident in subtype C conserved V3 loop, triplicate NF-B enhancers in the long terminal repeat region and mutations present in the Tat motif (Travers et al., 2004). A number of mutations have been identified and have shown a preference towards certain subtypes. The C30S31 mutation shows a greater propensity for subtype C HIV-1 strains while subtype B appears to be dominated by the C30C31 mutation. It has been suggested that the amino acid change, in the 31 position of the Tat motif is what makes subtype B so much more pathogenic that it's subtype C counterpart and aids in the progression of HAND by inducing apoptosis in neurons.

This study focused on infected patients in the Cape Town area and consisted of a small cohort of 64 patients. Previous studies conducted in South Africa has primarily focused on the clinical aspects of HAND such as the impact of HIV-1 on neurocognitive functioning and the effect of HAART in the CNS (Joska et al., 2012; Moshani 2011). This is the first South African study which attempts to investigate the viral epidemiology of a HAND patient cohort, identify signature pattern residues which are divergent between subtype C and subtype B sequences and analyze Tat motif mutations on a functional level.

Preliminary genetic subtyping of the HAND cohort identified 6 possible recombinant sequences with a large majority of them being identified as subtype C, which was expected as subtype C is one of the most prevalent subtypes in Cape Town (Engelbrecht et al., 2011). However, upon further inspection it was determined that a total of 15 sequences were in fact recombinant forms with either intra-subtypes or inter-subtypes being identified. The identification of inter-subtype recombinants not only indicates the introduction of new viral subtypes for the Cape Town HIV-1 population but also suggests a shift in the epidemiological paradigm. Furthermore, the identification of intra-subtype C recombinants suggests that the level of recombination in Cape Town may be higher than previously thought, as recombinant subtype C strains may mistakenly be identified as pure subtype C strains.

In addition 6 sequences were identified as having resistance mutations, which translates as 12.7% of the screened *pol* sequences. Given the fact that this is a small cohort it may be an inaccurate depiction of the actual level of resistance currently experienced in the study population. However, as resistance testing is only initiated after treatment failure, this finding may have implications for anti-retroviral treatment plans, especially since this is a treatment naïve cohort. This also suggests that treatment naïve patients with pre-existing resistance may have a greater chance of experiencing first line therapy failure.

Signature pattern analysis identified areas of relative conservation within the six domains of Tat between the three sequence subsets however this was outweighed by the level of variability observed in the cysteine rich, glutamine-rich and C-terminal domains. As expected, cohort sequences showed great similarity to other subtype C sequences from India and South Africa, although in some instances this similarity,

though to a lesser degree, was also shared between cohort sequences and subtype B sequences. In addition, while tat motif mutations in subtype B sequences were only C30C31, subtype C sequences from the cohort, South Africa and India had a mixture of both C30C31 and C30S31 mutations.

5.2 Strengths and Limitations of study for possible future work

This study has highlighted the importance of continued HIV-1 surveillance for the detection and proper identification of new circulating HIV-1 strains and the need for resistance screening of treatment naïve patients. However, a number of aspects in this study were found to be limiting and could be improved for possible future studies.

Firstly the cohort size was too small to give a clear indication of the level of diversity in the South African HAND study population. This was further compounded by the fact that cohort patients were only sampled from 3 clinics in Cape Town. The study however provided a foundation for further analyses which could possibly compare the diversity of HAND patients from various provinces.

Secondly, the lack of available data regarding Tat motif mutations in South Africa meant that any comparisons of the novel mutation (C30A31) identified in the cohort was restricted to subtype B and Indian subtype C sequences. The identification of a novel mutation has prompted further analyses and asks some important questions, what is the prevalence of the C30A31 mutation, is the C30A31 mutation genetically significant for subtype C, is there a possibility of other novel mutations circulating amongst the HIV infected and if so are they unique to the South African HAND population.

Thirdly, 293T cells, an adherent kidney cell line, were used for the functional analyses of HIV-1 Tat since it is easily cultured and transfected. However, HIV-1 Tat is neurotoxic in nature and it would therefore be useful to investigate the apoptotic effects of Tat in cell lines derived from neurons, monocytes and astrocytes as the latter are especially prone to infection by HIV-1 crossing the blood brain barrier into the CNS.

Finally, statistical analysis showed no significance in the level of apoptosis being induced by each clone. A possible explanation for this outcome could be that only three repeats were performed for this experiment. In addition the level of variability between each experiment with regard to cell passage number and cell viability may have influenced the overall level of apoptosis being detected. In order to overcome this problem and to clearly differentiate between typical cell apoptosis and Tat induced apoptosis, transfected cells could in future be labeled with a marker such as Halotag.

5.3 Conclusions

This project gave some important insight into HIV-1 *tat* in a HAND study cohort in South Africa; however larger studies are warranted to determine whether or not the identified trends are significant.

A high level of variability was detected with a total of 15 recombinants being identified. Possible recombinants consisted of 6 inter subtype recombinants and 9 intra-subtype C recombinants being identified. In addition 12.76% of *pol* sequences were identified with resistance mutations with half of these mutations manifesting in recombinants. Furthermore, signature pattern analysis identified variability between subtype C and subtype B *tat* sequences with the C30S31 dicysteine motif being relatively conserved in both Indian and South African subtype C sequences. However it should be noted that a number of sequences in the cohort and South African cohort also exhibited the wild type subtype B C30C31 mutation. In addition a novel mutation was identified in one sequence. This is the first identification of this mutation and may have implications for HAND as this mutation induced a similar level of apoptosis displayed by the C30C31 mutation.

However this project was not without its flaws, therefore the limitations previously discussed warrant further investigation before conclusions can be drawn about each Tat mutation and their ability to actively induce apoptosis in cell culture.

References

Amendt, B.A., Si Z.H., and Stoltzfus C.M. (1995). Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. *Molecular Cell Biology*. *15*:4606-46015.

Antinori, a, Arendt, G., Becker, J. T., Brew, B. J., Byrd, D. a, Cherner, M., Clifford, D. B., et al. (2007). Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*. 69:1789–99.

Ariën, K., Vanham, G., & Arts, E. (2007). Is HIV-1 evolving to a less virulent form in Humans? *Nature Reviews Microbiology*. *5:*141–151.

Arya, S., Guo, C., Josephs, S., & Wong-Staal, F. (1985). Trans-activator gene of human Tlymphotropic virus type III (HTLV-III). *Science*. *229*:69–73.

Ashkenazi, A. (2002). Targeting death and decoy receptors of the tumour-necrosis factor superfamily. *Nature reviews:Cancer*. 2:420–30.

Ataher, Q., Portsmouth, S., Napolitano, L. A, Eng, S., Greenacre, A., Kambugu, A., Wood, R., Badal-Faesen, S., and Tressler, R. (2012). The epidemiology and clinical correlates of HIV-1 co-receptor tropism in non-subtype B infections from India, Uganda and South Africa. *Journal of the International AIDS Society*. *15*:2-10.

Badley, A. (2005). In vitro and in vivo effects of HIV Protease inhibitors on apoptosis. *Cell death and differentiation*, *12*:924–931.

Badley, A., Roumier, T., Lum, J., & Kroemer, G. (2003). Mitochondrion-mediated apoptosis in HIV-1 infection. *Trends in Pharmacological Science*. *24*:298–305.

Badou, A., Bennasser, Y., Moreau, M., Leclerc, C., Benkirane, M., & Bahraouhi, E. (2000). Tat Protein of Human Immunodeficiency Virus Type 1 Induces Interleukin-10 in Human Peripheral Blood Monocytes: Implication of Protein Kinase C-Dependent Pathway. *Journal of Virology*. *74*:10551–10562.

Ball, S., Abraha, A., Collins, K., Marozsan, A. J., Baird, H., Quinones-Mateu, M. E., Penn-Nicholson, A., Murray,M., Richard, N., Lobritz,M., Zimmerman,P. A., Kawamura, T., Blauvelt, A., & Arts, E. J. (2003). Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *Journal of virology*. *77*: 1021-1038.

Barré-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, F., Vezinet-Brun, C., Rozenbaum, W., & Montagnier, L.,(1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. *220*:868–871.

Basavapathruni, A., & Anderson, K. S. (2007). Reverse transcription of the HIV-1 pandemic. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. *21*:3795–808.

Becker, M., Spracklen, F., & Becker, W. (1985). Isolation of a Lymphadenopathy associated virus from a patient with acquired immune deficiency syndrome. *South African Medical Journal*. *68*:144–147.

Blazek R., Schmitt K., Krafft, U., & Hadding, U. (1990). Fast and simple procedure for the
detection of cell culture mycoplasmas using a single monoclonal antibody. J*ournal of Immunological Methods*. *131*: 203-212.

Boissé, L., Gill, M., & Power, C. (2008). HIV infection of the central nervous system: clinical features and neuropathogenesis. *Neurologic Clinics*. *26*:799–819.

Bouillet, P., & Strasser, A. (2002). BH3-only proteins – evolutionarily conserved pro-apoptotic Bcl-2 family members essential for initiating programmed cell death. *Journal of cell science*. *115*:1567–1574.

Bredell H., Hunt G., Casterling A., Cilliers T., Rademeyer C., Coetzer M., Miller S., Johnson D., Williamson C., and Morris L. (2002). HIV-1 Subtype A, D, G, AG and Unclassified Sequences identified in South Africa. *Aids Research and Human Retroviruses*. *18*:681-683.

Brew, B. (2001). Markers of AIDS dementia complex: the role of cerebrospinal fluid assays. *AIDS*. *15*:1883–1884.

Briggs, J. A. G., Wilk, T., Welker, R., Kräusslich, H-G., & Fuller, S. D. (2003). Structural organization of authentic, mature HIV-1 virions and cores. *The EMBO journal*. *22*: 1707–1715.

Brouwers, P., Hendricks, M., Lietzau, J., Pluda, J., Mitsuya, H., Broder, S., & Yarchoan, R. (1997). Effect of combination therapy with zidovudine and didanosine on neuropsychological functioning in patients with symptomatic HIV disease: a comparison of simultaneous and alternating regimens. *AIDS*. *11*:59–66.

Buscemi, L., Ramonet, D., & Geiger, J. D. (2007). Human immunodeficiency virus type-1 protein Tat induces tumor necrosis factor-alpha-mediated neurotoxicity. *Neurobiology of disease*. *26*:661–70.

Campbell, G. R., & Loret, E. P. (2009). What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? *Retrovirology*. 6:50-62.

Campbell, G., Watkins, J., Esquieu, D., Pasquier, E., Loret, E., & Spector, S. (2005). The C terminus of HIV-1 Tat modulates the extent of CD178-mediated apoptosis of T cells. *Journal of Biochemical Chemistry*. *280*:38376–38382.

Cann, A., Rosenblatt, J., Wachsman, W., Shah, N. P., & Chen, I. (1985). Identification of the gene responsible for human T-cell leukaemia virus transcriptional regulation. *Nature*. *318*:571–574.

Chan, D., Fass, D., Berger, J., & Kim, P. (1997). Core structure of gp41 from the HIV Envelope Glycoprotein. *Cell*. *89*:263–273.

Clavel, F., & Hance, A. J. (2004). HIV Drug Resistance. New England Journal of Medicine. *350*:1023-1035.

Clever, J., & Parslow, T. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *Journal of Virology*. 71:3407–3414.

Contreras, X., Bennasser, Y., Chazal, N., Moreau, M., Leclerc, C., Tkaczuk, J., & Bahraouhi, E. (2005). Human immunodeficiency virus type 1 Tat protein induces an intracellular calcium increase in human monocytes that require DHP receptors: involvement in TNF-alpha production. *Virology*. *332*:316–328.

Cowley, D., Gray, L. R., Wesselingh, S. L., Gorry, P. R., & Churchill, M. J. (2010). Genetic and functional heterogeneity of CNS-derived tat alleles from patients with HIV-associated dementia. *Journal of neurovirology*. *17*:1–12.

Das, A., Klaver, B., & Berkhout, B. (1998). The 5′ and 3′ TAR elements of human immunodeficiency virus exert effects at several points in the virus life cycle. *Journal of Virology*. *72*:9217–9223.

Dey, S. S., Xue, Y., Joachimiak, M. P., Friedland, G. D., Burnett, J. C., Zhou, Q., Arkin, A. P., & Schaffer, D. A. (2012). Mutual Information Analysis Reveals Coevolving Residues in Tat That Compensate for Two Distinct Functions in HIV-1 Gene Expression. *The Journal of Biological Chemistry.* 287:7945–7955.

de Oliveira, T., Engelbrecht, S., Janse van Rensburg, Gordon, E. M., Bishop, K., zur Magede, J., Barnett, S., & Cassol, S. (2003). Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness? *Journal of Virology*. *77*:9422–9430.

Denault, J-B., & Salvesen, G. S. (2002). Caspases: keys in the ignition of cell death. *Chemical reviews. 102*:4489–500.

Desfosses, Y., Solis, M., Sun, Q., Grandvaux, N., Van Lint, C., Burny, A., Gatignol, A., et al. (2005). Regulation of human immunodeficiency virus type 1 gene expression by clade-specific Tat proteins. *Journal of Virology*. *79*:9180– 9191.

Dicker, I., Samantha, H., Li, Z., Hong, Y., Tian, Y., Banville, J., Remillard, R., et al. (2007). Changes to the HIV Long Terminal repeat and to HIV Integrase Differentially Impact HIV Integrase Assembly, Activity, and the Binding of Strand Transfer Inhibitors. *Journal of Biological Chemistry. 43*:31186–1196.

Engelbrecht, S., Laten, J., Smith, T-L., & van Rensburg, E. J. (1995). Identification of env subtypes in fourteen HIV type 1 isolates from South Africa. *AIDS Research and Human Retroviruses*. *11*:1269–1271.

Engelbrecht, S., Wilkinson, E., de Oliveira, T. (2011)Diversity and molecular epidemiology of HIV-1 in Cape Town; 1984 to 2010. Retroviriology. *8*:21.

Evans, P., Dampier, W., Ungar, L., & Tozeren, A. (2009). Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Medical Genomics. 2*:27-39.

Finzi, D., & Siliciano, R. (1998). Viral dynamics in HIV-1 infection. *Cell*. *93*:665–671.

Fittipaldi, A., Ferrari, A., Zoppe, M., Arcangeli, C., Pellegrini, V., Beltram, F., & Giacca, M. (2003). Cell membrane lipid rafts mediate caveolar endocytosis of HIV-1 Tat fusion proteins. *Journal of Biochemical Chemistry*. *278*:34141–341349.

Frankel, D., & Young, J. (1998). HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry*. *67*:1–25.

Follansbee, S. E., Busch, D. F., Wofsy, C. B., Coleman, D. L., Gullet, J., Aurigemma, G. P., Ross, T., Hadley, W. K., Drew, W. L. (1982). Annals of Internal Medicine. *96*:705-713.

Annals of Internal Medicine [1982, 96(6 Pt 1):705-713]

Freed, E. O. (2001). HIV-1 replication. *Somatic cell and molecular genetics*, *26*(1-6), 13–33.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., et al. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*. *397*:436–41.

Gifford, R.J., Liu T.F., Rhee S.Y., Kiuchi M., Hue S., Pillay D., and Shafer R.W. (2009). The calibrated population resistance tool: standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics.*

Gisslén, M., Norkrans, G., Svennerholm, B., & Hagberg, L. (1997). The effect on human immunodeficiency virus type 1 RNA levels in cerebrospinal fluid after initiation of zidovudine or didanosine. *Journal of Infectious Disease*. *175*:434–437.

Gottlieb, M., Schroff, R., Schanker, H., Weisman, J., Fan, P., Wolf, R., & Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: Evidence of a new acquired cellular immune deficiency. *New England Journal of Medicine*. *305*:1425–1428.

GLOBAL AIDS RESPONSE REPUBLIC OF SOUTH AFRICA (2012) Report *12*.

Grant, I. (2008). Neurocognitive disturbances in HIV. *International review of psychiatry. 20*:33–47.

Gupta, J. D., Satishchandra, P., Gopukumar, K., Wilkie, F., Waldrop-Valverde, D., Ellis, R., Ownby, R., et al. (2007). Neuropsychological deficits in human immunodeficiency virus type 1 clade C-seropositive adults from South India. *Journal of neurovirology*. *13*:195–202.

Gutheil, W. G., Subramanyam, M., Flentke, G., Sanford, D., Munoz, E., Huber, B., & Bachovchin, W. (1994). Human immunodeficiency virus 1 Tat binds to dipeptidyl aminopeptidase IV (CD26): a possible mechanism for Tat's immunosuppressive activity. *Proceedings of the National Academy of Sciences*. *91*:6594–6598.

http://www.hiv.lanl.gov/ - HIV Sequence Compendium. Leitner T, McCutchan F, Foley B, Mellors JW, Hahn B, Wolinsky S, Marx P and Korber B. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, LA-UR 03-3564

Jacobs, G.B., Loxton, A.G., Laten, A., Robson, B., Janse van Rensburg, E., & Engelbrecht, S. (2009). Emergence and Diversity of Different HIV-1Subtypes in South Africa, 2000–2001. *Journal of Medical Virology*. *81*:1852–1859.

Jacobs, G. B., Laten A. D., van Rensburg E. J., Bodem J., Weissbrich B., Rethwilm A., Preiser W., & Engelbrecht S. (2008). Phylogenetic Diversity and Low Level Antiretroviral Resistance Mutations in HIV Type 1 Treatment-Naive Patients from Cape Town, South Africa. *AIDS Research and Human Retroviruses*. *24*:1009-1012.

Jeang, K. (1996). HIV-1 Tat: structure and function. *sequence compendium*. Retrieved from http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/1996/PART-III/1.

Jeang, K. T., Xiao, H., & Rich, E. A. (1999). Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *The Journal of biological chemistry*. *274*:28837–28840.

Jevtović, D., Vanovac, V., Veselinović, M., Salemović, D., Ranin, J., & Stefanova, E. (2009). The incidence of and risk factors for HIV-associated cognitive-motor complex among patients on HAART. *Biomedical Phamacotherapy. 63*:561–565.

Joska, J. A, Fincham, D. S., Stein, D. J., Paul, R. H., & Seedat, S. (2010). Clinical correlates of HIV-associated neurocognitive disorders in South Africa. *AIDS and behavior*. *14*: 371–8.

Kamla V., Henrich B., & Hadding U. (1992). Species differentiation of mycoplasmas by EFTu specific monoclonal anitbodies. *Journal of Immunological Methods*. *147*:73-81.

Karn, J. (1999). Tackling Tat. *Journal of molecular biology*. *293*:235–54.

Kelly, G., Ensoli, B., Gunthel, C., & Offermann, M. (1998). Purified Tat induces inflammatory response genes in Kaposi's sarcoma cells. *AIDS*. *12*:1753–1761.

Kiebala, M. M. (2010). *Targeting HIV-1 Associated Neurocognitive Disorder (HAND): Endogenous and Exogenous Anti-Inflamatory Mechanisms*.

Kim, H., & Perelson, A. S. (2006). Viral and latent reservoir persistence in HIV-1-infected patients on therapy. *PLoS computational biology. 2*:135.

Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. *16*:111-120.

Khoja, S., Ojwang, P., Khan, S., Okinda, N., Harania, R., & Ali, S. Genetic Analysis of (2008). HIV-1 Subtypes in Nairobi, Kenya. *PLoS one*. *3*: e3191.

Korber, B., & Myers G. Signature pattern analysis: a method for assessing viral sequence relatedness. (1992). *AIDS Research Human Retroviruses*. *8*:1549-60.

Lalonde, M. S., Lobritz, M. a, Ratcliff, A., Chamanian, M., Athanassiou, Z., Tyagi, M., Wong, J., et al. (2011). Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tat-transactivating response element (TAR) RNA. *PLoS pathogens*. *7*:e1002038.

Li,G-H., Li,W., Mumper,R. J., & Nath, A. (2012). Molecular mechanisms in the dramatic enhancement of HIV-1 Tat transduction by cationic liposomes. *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology. doi: 10.1096/fj.11-203315.*

Li, C., Wang, C., Friedman, D., & Pardee, A. (1995). Reciprocal modulations between p53 and Tat of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences*. *92*:5461–5464.

Lin, J., Chiba, M., Balani, S., Chen, I., Kwei, G., Vastag, K., & Nishime, J. (1996). Species differences in the pharmacokinetics and metabolism of indinavir, a potent human immunodeficiency virus protease inhibitor. *Drug Metabolism and Disposition*. *24*: 1111–1120.

Liner, K. J., Hall, C. D., & Robertson, K. R. (2007). Impact of human immunodeficiency virus (HIV) subtypes on HIV-associated neurological disease. *Journal of neurovirology*. *13*:291–304.

Loxton AG, Treurnicht F, Laten A, Janse Van Rensburg E, and Engelbrecht S. Sequence analysis of near full-length HIV type 1 subtype D primary strains 130 isolated in Cape Town, South Africa, from 1984 to 1986. (2005). *AIDS Research and Human Retroviruses*. 21:410-413

Lu, K., Heng, X., & Summers, M. F. (2011). Structural determinants and mechanism of HIV-1 genome packaging. *Journal of molecular biology*. *410*:609–633.

Meier, P., Finch, A, & Evan, G. (2000). Apoptosis in development. *Nature*. *407*:796–801.

Mishra, M., Vetrivel, S., Siddappa, N. B., Ranga, U., & Seth, P. (2008). Clade-specific differences in neurotoxicity of human immunodeficiency virus-1 B and C Tat of human neurons: significance of dicysteine C30C31 motif. *Annals of neurology*, *63*(3), 366–76.

Nisole, S., & Saïb, A. (2004). Early steps of retrovirus replicative cycle. *Retrovirology*. *1*:9.

O'Connor, L., Huang, D.C.S., O'Reilly, L. A., and Strasser, A. (2000). Apoptosis and cell division. *Current Opinion in Cell Biology*. 257–263.

Odiase, F., Ogunrin, O., & Ogunniyi, A. (2006). Effect of progression of disease on cognitive performance in HIV/AIDS. *Journal of the National Medical Association*. *98:*1260–1262.

Osmanov, S., Pattou, C., Walker, N., Schwardländer, B., & Esparza, J. (2002). Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *Journal of acquired immune deficiency syndromes (1999)*. *29*:184–90.

Papathanasopoulos, MA Cilliers, T., Morris, L., Mokili, J., Dowling, W., Brix, D., & McCutchan, F. (2002). Full-Length Genome Analysis of HIV-1 Subtype C Utilizing CXCR4 and Intersubtype Recombinants Isolated in South Africa. *AIDS Research and Human Retroviruses*. *18*:879–886.

Parsons, T. D., Braaten, A. J., Hall, C. D., & Robertson, K. R. (2006). Better quality of life with neuropsychological improvement on HAART. *Health and quality of life outcomes*. *4*:1477-7525

Peña, A. C. P., Faria, N. R., Imbrechts , S., Libin, P., Abecasis, A. B., Deforche, K., Gomez A., Camacho R. J., de Oliveira T., Vandamme A-M. Performance of the Subtyping Tools in the Surveillance of HIV-1 Epidemic: Comparison Between Rega Version 3 and Six Other Automated Tools to Identify Pure Subtypes and Circulating Recombinant Forms. 17th International BioInformatics Workshop on Virus Evolution and Molecular Epidemiology University of Belgrade, Faculty of Medicine, Belgrade, Serbia, 27 - 31 August 2012. Abstract presentation.

Piot, P., Ouinn, T., Taelman, H., Feinsod, F., Minlangu, K., Wobin, O., Mbendi, N., et al. (1984). Acquired immunodeficiency syndrome in a heterosexual population in Zaire. *Lancet*, *2*:65–69.

Plantier, J-C., Leoz, M., Dickerson, J. E., de Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D. L., & Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine*. *15*:871 - 872.

Plantier, J-C., Dachraoui R., Lemee V., Gueudin M., Borsa-Lebas F., Caron, F., & Simon, F. HIV-1 resistance genotyping on dried serum spots. (2005). *AIDS*. *19*:391-397.

Pond, S. L., and Smith, D. M. (2009). Are All Subtypes Created Equal? The Effectiveness of Antiretroviral Therapy against Non–Subtype B HIV-1. Clinical Infectious Diseases. 48:1306–1309.

Pond, S. L., Posada, D., Stawiski, E., Chappey, C., Poon, A.F.Y., Hughes, G., Fearnhill, E., Gravenor, M. B., Brown, A. J.L and Frost, S.D.W. (2009). An Evolutionary Model-Based Algorithm for Accurate Phylogenetic Breakpoint Mapping and Subtype Prediction in HIV-1.PLoS Computational Biology. 5:e1000581.

Pu, H., Tian, J., Flora, G., Woo Lee, Y., Nath, A., Hennig, B., & Toborek, M. (2003). HIV-1 tat protein upregulates inflammatory mediators and induces monocyte invasion into the brain. *Molecular and Cellular Neuroscience*. *24*:224–237.

Pugliese, A., Vidotto, V., Beltramo, T., Petrini, S., & Torre, D. (2005). A review of HIV-1 Tat protein biological effects. *Cell biochemistry and function. 23*:223–227.

Ragupathy, V., Zhao, J., Wood, O., Tang, S., Lee, S., Nyambi, P., & Hewlett, I. (2011). Identification of new, emerging HIV-1 unique recombinant forms and drug resistant viruses circulating in Cameroon. *Virology Journal*. *8*:185.

Ranga, U., Shankarappa, R., Siddappa, N., Ramakrishna, L., Nagendran, R., Mahalingam, M., Mahadevan, A., et al. (2004). Tat Protein of Human Immunodeficiency Virus Type 1 Subtype C Strains Is a Defective Chemokine. *Journal of Virology*. *78*:2586–2590.

Rao, V. R., Sas, A. R., Eugenin, E. A., Siddappa, N. B., Bimonte-nelson, H., Berman, J. W., Ranga, U., et al. (2008). HIV-1 clade specific differences in the induction of neuropathogenesis. *Journal of neuroscience*. *28*:10010–10016.

Ras, G., Simson, I., Anderson, R., Prozeksy, O., & Hamersma, T. (1983). Acquired immunodeficiency syndrome: a report of 2 South African cases. *South African Medical Journal*. *64*:140–142.

Rayne, F., Vendeville, A., Bonhoure, A., & Beaumelle, B. (2004). The ability of chloroquine to prevent tat-induced cytokine secretion by monocytes is implicated in its in vivo anti-human immunodeficiency virus type 1 activity. *Journal of Virology*. *78*:12054–12057.

Rhee S., Gonzales, M. J., Kantor, R., Betts, B. J., Ravela, J & Shafer, R. W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research.* 31:298-303.

Romani, B., Engelbrecht, S., & Glashoff, R. H. (2010). Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *The Journal of general virology*. *91:*1–12.

Romani, B. (2010). Mutagenesis and functional studies of the HIV-1 vpr gene and Vpr protein obtained from South African virus strains. PhD thesis. Stellenbosch University.

Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J.N., Lee, T. C., McLauglin, S., Moore, S. W., Shindo, N., Learn, G.H., Korber, B.T., Brander, C., Goulder, P. J. R. , Kiepiela, P., Walker, B. D., & Mullins, J. I. (2006). Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *Journal of Virological Methods*. *136*:118-125.

Roshal, M., Zhu, Y., & Planelles, V. (2001). Apoptosis in AIDS. *Apoptosis*. *6*:103–116.

Sacktor, N. C., Wong, M., Nakasujja, N., Skolasky, R. L., Selnes, O. A, Musisi, S., Robertson, K., et al. (2005). The International HIV Dementia Scale: a new rapid screening test for HIV dementia. *AIDS.19*:1367–74.

Sacktor, N., Nakasujja, N., Skolasky, R., Robertson, K., Wong, M., Musisi, S., Ronald, A, et al. (2006). Antiretroviral therapy improves cognitive impairment in HIV+ individuals in sub-Saharan Africa. *Neurology*. *67*:311–4.

Sawaya, B., Thatikunta, P., Denisova, L., Brady, J., Khalili, K., & S, A. (1998). Regulation of TNFalpha and TGFbeta-1 gene transcription by HIV-1 Tat in CNS cells. *Journal of Neuroimmunology*. *87*:33–42.

Schultz, A. K., Bulla, I., Abdou-Chekaraou, M., Gordien, E., Morgenstern, B., Zoulim, F., Dény, P., Stanke, M. (2012). jpHMM: recombination analysis in viruses with circular genomes such as the hepatitis B virus. *Nucleic Acids Research*. *40*:W193-W198.

Schultz, A., Zhang, M., Leitner, T., Kuiken, C., Korber, B., Morgenstern, B., & Stanke, M. (2009). A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes. *BMC Bioinformatics*. *7*:265–280.

Shafer, R. W., Rhee, S-R., Pillay, D., Miller, V., Sandstrom, P., Schapiro, J. M., Kuritzkes, D. R., & Bennett, D. (2007). HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS*. *21*:215–223.

Sher, R. (1989). HIV infection in South Africa, 1982 - 1988 - a review. *South African Medical Journal*. *76*:314–318.

Soares, E. A., Santos, A. F., Gonzalez, L. M., Lalonde,M. S., Tebit, D. M., Tanuri, A., Arts E. J., & Soares, M. A. (2009). Mutation T74S in HIV-1 subtype B and C proteases resensitizes them to ritonavir and indinavir and confers fitness advantage. J*ournal of Antimicrobial Chemotherapy*. *64*:938–944.

Strasser, A., O'Connor, L., & Dixit, V. (2000). Apoptosis signaling. *Annual review of biochemistry*, *69*, 217–245.

Suzuki, Y., & Suzuki, Y. (2005). Gene Regulatable Lentiviral Vector System. *Viral gene therapy*. Ke Xu (Ed.), ISBN: 978-953-307-539-6, InTech, Available from: http://www.intechopen.com/books/viral-gene-therapy/gene-regulatable-lentiviral-vector-system.

Swanson P., Devare, S. G., & Hackettt J. Jr. (2003). Molecular Characterization of 39 HIV-1 Isolates Representing Group M (Subtype A-G) and Group O: Sequence Analysis of gag p24, pol Integrase, and env gp41. *AIDS Research and Human Retroviruses*. *19*:625-629

Tebit, D. M., & Arts, E. J. (2011). Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *The Lancet infectious diseases*. *11*:45–56.

Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., & Higgins, D. (1997). The CLUSTAL X windows interface: Flexible strategies for multiple-sequence alignment aided by quality analysis tools. *Nucleic Acids Research. 25*:4876–4882.

Tiwari, S., Nair, M. P.N., and Shailendra, K. S. (2012). Rendezvous with Tat: Transactivator of Transcriptionduring Human Immunodeficiency Virus Pathogenesis. *American Journal of Infectious Disease.* 8:79-91.

Tozzi, V., Balestra, P., Bellagamba, R., Corpolongo, A., Salvatori, M. F., Visco-comandini, U., Vlassi, C., et al. (2007). Persistence of Neuropsychologic Deficits Despite Long-Term Highly Active Antiretroviral Therapy in Patients with HIV-related Neurocogntive Impairment Prevalence and Risk Factors. *45*:174–182.

Travers, S. A. A., Clewley, J. P., Glynn, J. R., Fine, P. E. M., Crampin, A. C., Sibande, F., Mulawa, D., et al. (2004). Timing and Reconstruction of the Most Recent Common Ancestor of the Subtype C Clade of Human Immunodeficiency Virus Type 1. *Journal of Virology*. *78*:10501–10506.

Tyagi, M., Rusnati, M., Presta, M., & Giacca, M. (2001). Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. *The Journal of Biological Chemistry*. *276*: 3254–61.

UNAIDS 2008. 2008 Report on the Global AIDS epidemic. - http://www.unaids.org/en/

Valcour, V. G., Sithinamsuwan, P., Nidhinandana, S., Thitivichianlert, R. K., Apateerapong, S., Shiramizu, W. B., Desouza, M., Chitpatima, S., Watt, G., Chuenchitra, T., et al. (2007). Neuropsychological abnormalities in patients with dementia in CRF 01 AE HIV-1 infection. *Neurology*. *68*:525–527.

van Harmelen J., Wood R., Lambrick M., Rybicki E. P., Williamson A. L., & Williamson C. (1997) An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS*. *11*:81-87.

Vendeville, A., Rayne, F., Bonhoure, A., Bettache, N., Montcourrier, P., & Beaumelle, B. (2004). HIV-1 Tat enters T cells using coated pits before translocating from acidified endosomes and eliciting biological responses. *Molecular Biology of the Cell*. *15*:2347–2360

Vomelova, I., Z. Vanickova, Z., & Sedo, A. (2009). Methods of RNA Purification. All Ways (Should) Lead to Rome. *Folia Biologica (Praha)*. *251*:243–251.

Westerndorp, M., Frank, R., Ochsenbauer, C., Stricker, K., Dhein, J., Walczak, H., Debatin, K., et al. (1995). Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. *Nature*. *375*:497–500.

Williamson, C., Engelbrecht, S., Lambrick, M., Janse van Rensburg, E Wood, R., Bredell, W., & Williamson, A. (1995). HIV-1 subtypes in different risk groups in South Africa. *Lancet*. *346*:782.

Xiao, H., Neuveut, C., Tiffany, H., Benkirane, M., Rich, E., Murphy, P., & Jeang, K. (2000). Selective CXCR4 antagonism by Tat: Implications for in vivo expansion of coreceptor use by HIV-1. *Proceedings of the National Academy of Sciences*. *97*:11466–11471.

Zauli, G., Gibellini, D., Celeghini, C., Mischiati, C., Bassini, A., La Placa, M., & Capitani, S. (1996). Pleiotropic effects of immobilized versus soluble recombinant HIV-1 Tat protein on CD3-mediated activation, induction of apoptosis, and HIV-1 long terminal repeat transactivation in purified CD4+ T lymphocytes. *Journal of Immunology*. *157*:2216–2224.

Zhang, M., Schultz, A., Calef, C., Kuiken, C., Leitner, T., Korber, B., Morgenstern, B., et al. (2006). jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. *Nucleic Acid Research*, *34*, 463–465.

Appendix A

Table 1: Patient demographics

Sample #	Gender	Age	Clinic	CD4 count cells/µl	Clinical
JO152	Female	30	WDH	185	MND
JO161	Female	32	NHC	307	MND
JO162	Female	26	NHC	361	MND
JO166	Female	33	UBN	75	Dementia
JO167	Female	35	NHC	409	MND
JO168	Female	30	UBN	335	MND
JO172	Female	29	NHC	929	Dementia
JO173	Female	34	NHC	398	MND
JO176	Female	26	NHC	282	Normal
JO179	Female	25	UBN	143	MND
JO180	Female	32	WDH	184	Normal
JO183	Female	33	UBN	201	MND
JO185	Female	25	NHC	318	MND
JO186	Female	31	UBN	152	ANI
JO189	Male	32	UBN	194	MND
JO190	Female	30	UBN	217	MND
JO191	Male	29	UBN	112	ANI
JO192	Female	35	NHC	172	MND
JO193	Male	26	UBN	172	MND
JO195	Male	35	UBN	146	ANI
JO196	Male	20	UBN	116	NA
JO198	Female	26	UBN	140	Dementia
JO200	Male	34	WDH	NA	Normal
JO201	Male	30	WDH	NA	MND
JO202	Female	28	UBN	107	Dementia
JO203	Female	29	UBN	180	Normal
JO204	Female	34	UBN	199	Normal
JO206	Female	30	NHC	198	Dementia
JO207	Female	35	NHC	192	MND
JO209	Female	33	WDH	118	MND
JO210	Female	33	UBN	120	Normal

JO213	Male	27	UBN	164	Normal
JO214	Female	28	WDH	165	MND
JO215	Female	32	WDH	192	Normal
JO216	Male	35	WDH	107	Dementia
JO218	Female	25	UBN	191	NA
JO219	Female	36	UBN	185	NA
JO220	Male	26	UBN	196	MND
JO223	Female	31	WDH	190	Normal
JO224	Female	35	WDH	182	Dementia
JO225	Female	34	UBN	196	Dementia
JO229	Female	29	UBN	180	Dementia
JO230	Female	32	UBN	100	Normal
JO231	Female	27	UBN	190	Normal
JO232	Male	32	UBN	135	Normal
JO233	Female	30	UBN	107	ANI
JO234	Male	30	Unknown	81	Normal
JO235	Female	30	WDH	NA	Normal
JO236	Female	32	UBN	139	Normal
JO237	Female	31	UBN	184	MND
JO238	Male	32	UBN	71	MND
JO239	Female	32	UBN	191	MND
JO240	Female	31	UBN	177	Dementia
JO241	Male	31	UBN	52	MND
JO244	Female	28	UBN	66	MND
JO247	Female	33	NHC	184	MND
JO248	Male	25	NHC	118	Normal
JO252	Female	25	WDH	186	Dementia
JO254	Female	27	UBN	182	MND
JO258	Male	25	UBN	168	Normal
JO259	Male	21	NHC	NA	NA
JO288	Female	Unknown	UBN	140	NA